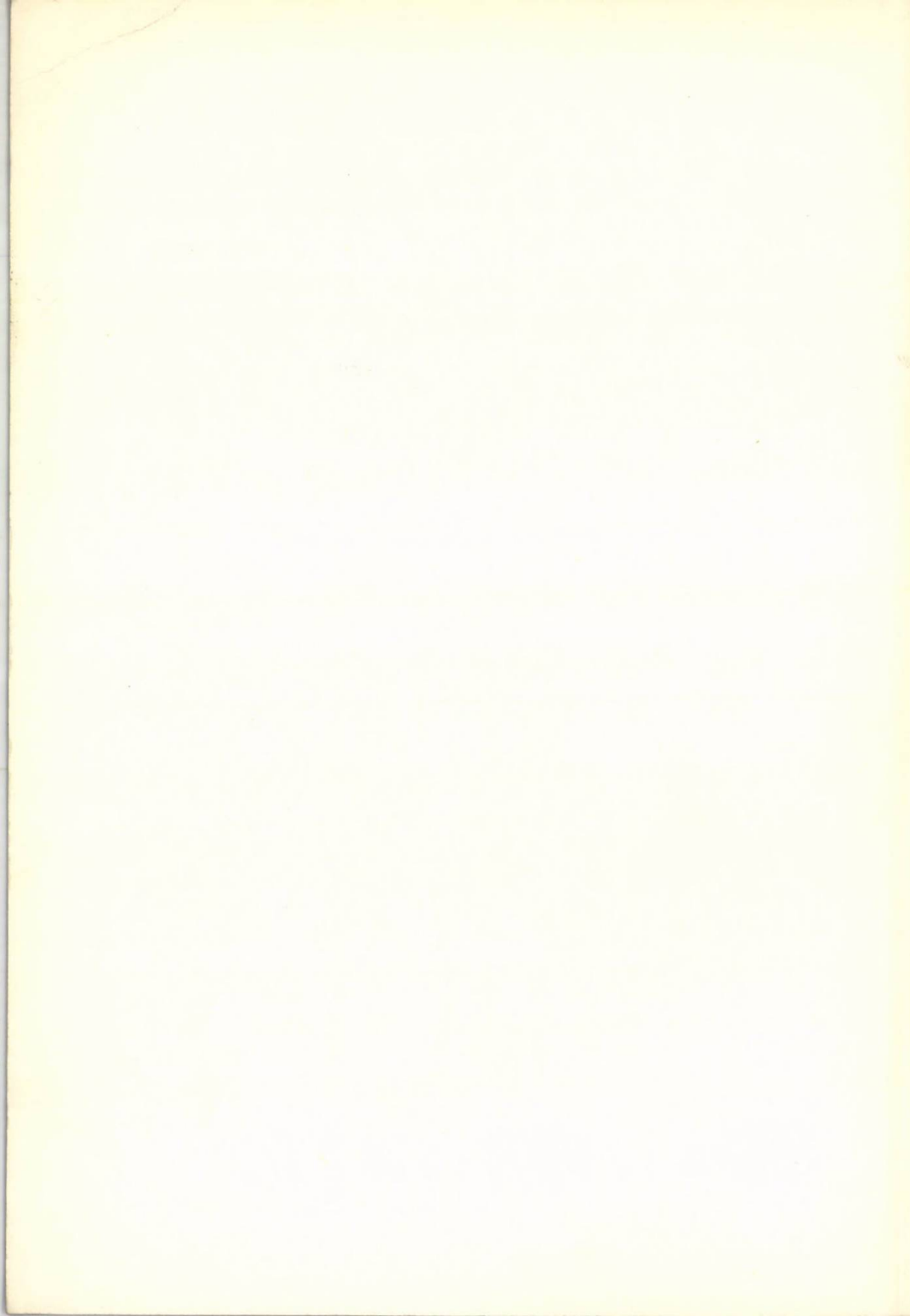


ALEXANDRA
ROGERIC
ALEXANDRESCU

**FONO-
STATISTICA
LIMBII
ROMÂNE**



**FONOSTATISTICA
LIMBII ROMÂNE**

Coperta de: *Negrescu Dumitru*

4
R 63

ALEXANDRA ROCERIC-ALEXANDRESCU

FONOSTATISTICA LIMBII ROMÂNE



EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA

1968

NOV 21 1941
U.S. AIR FORCE



0. INTRODUCERE

0.1. În cele ce urmează ne propunem să descriem structura fonologică a limbii române din punct de vedere cantitativ. Descrierea noastră selectează din *limba română standard* numai aspectul limbă română contemporană.

Metoda de cercetare pe care o prezentăm s-a constituit ca metodă științifică de lucru destul de recent, aparținând lingvisticii moderne. De aceea, am considerat utile o serie de observații preliminare, de introducere în acest domeniu.



0.2. Până nu de mult, lingvistica matematică era singurul termen general prin care erau denumite toate cercetările avînd ca obiect de studiu limba, iar ca metodă, matematica. Lingvistica matematică — în ansamblu — este o disciplină nouă. Din domeniul ei vast se desprind o serie de direcții care ne oferă, fiecare, imagini specifice asupra limbii. Nu este cazul să intrăm în amănunte privind toate disciplinele pe care le cuprinde lingvistica matematică¹. Ne oprim la una dintre ele: *statistica*.

Lingvistica statistică este acea ramură a lingvisticii care stabilește raporturi statistice între elementele unui idiom, la un moment dat. Aceste mărimi sînt exprimate numeric prin frecvențe relative. Se adaugă astfel, la criteriile: posibilitate de apariție a unui element într-o situație dată și relații formale între tipuri de structuri, caracterizarea printr-o anumită măsură considerată relevantă.

¹ De altfel, s-a scris destul de mult pe această temă. În lingvistica românească, există, printre altele, o expunere cuprinzătoare: Solomon Marcus, Edmond Nicolau, Sorin Stati, *Introducere în lingvistica matematică*.

Spre deosebire de alte tipuri de investigație, statistica pune în lumină faptele care reprezintă *media* în limbă, tendințele, variabilele și relațiile, în general, lăsând de o parte detaliile. Ea ne informează deci asupra orientării generale și nu asupra evenimentelor particulare. Metoda structurală calitativă cercetează și ea, de fapt, unele aspecte globale din limbă, dar fără a preciza vreun fel de indici numerici. Avînd asemenea obiective, metoda statistică este deosebit de utilă în analiza unui ansamblu cu multe elemente. De aici eficacitatea ei în ce privește studiile de limbă. Se știe că, în comunicare, oamenii folosesc un număr foarte mare de simboluri încît uneori avem impresia că numărul de semne este nelimitat. „Simplificarea” statistică ne dă posibilitatea de a sistematiza aceste elemente și de a observa că nu tot ce se comunică (se vorbește, se scrie) este nou, irepetabil. Dimpotrivă, există secvențe care revin pentru că sistemul limbii exercită o anumită constrîngere, limitînd posibilitățile de a face combinații de semne. Altfel spus, metoda statistică permite să se stabilească tocmai acele regularități cu caracter global care nu se pot detecta prin alte metode. Este evident, în același timp, că efectuarea unei cercetări cantitative în limbă nu înseamnă și nu poate însemna o investigație completă și nu este în măsură să illustreze integral un material, datorită fenomenului atît de complex pe care îl reprezintă limba.

Semnele lingvistice — fie ele sunete, litere etc. — se repetă cu o frecvență fixă într-o limbă dată, limba avînd un caracter ergodic. Așa se enunță postulatul pe care se bazează aplicarea metodei statistice. Aceasta este de altfel și un fapt observat și verificat universal, încît poate fi considerat ca o lege lingvistică².

Prin metoda ei, cercetarea științifică reclamă, în general, două tipuri de analiză: calitativă și cantitativă. Analiza calitativă observă și clasifică tipurile de fenomene după natura lor, pe cînd analiza cantitativă (în care se include și analiza statistică) apreciază mărimea și frecvența lor. Evident, judecata calitativă o precede pe cea cantitativă, care-i furnizează date privitoare la *ce* sînt, *ce* reprezintă elementele care urmează să fie măsurate. Ca și în celelalte științe, în lingvistică se aplică ambele tipuri de cercetare, metoda fiind aceeași, diferind numai obiectul.

Concepînd lingvistica în ansamblul celorlalte științe, înțelegem mai ușor de ce o cercetare calitativă pură este incompletă, proporția elementelor din limba descrisă rezultînd numai în urma aplicării

² Pierre Guiraud, *Problèmes et méthodes de la statistique linguistique*, p. 16.

analizei cantitative³. La aceasta trebuie să adăugăm că statistica nu poate fi utilă lingvisticii decât dacă e întrebuințată ca un instrument de lucru, ea rămânând subordonată obiectului esențial, care este analiza calitativă, structurală a limbii (chiar dacă, istoricește, elemente de statistică lingvistică precedă uneori lingvistica structurală). De aceea înainte de a începe orice fel de calcul, trebuie să stabilim ce probleme lingvistice urmează să clarificăm pe această cale. Prin determinarea caracteristicilor cantitative ale unor fenomene de limbă, nu înțelegem în nici un caz introducerea lor obligatorie în limitele unor formule matematice, sau „deducerea” lor din structuri matematice enunțate anterior.

Există însă moduri deosebite de a concepe lingvistica statistică.

a) O serie de cercetări abordează fenomenele lingvistice făcînd apel la tehnici matematice, pentru a trage concluzii sau a pune în evidență trăsături ale limbii care nu s-ar putea degaja printr-un studiu pur calitativ. În acest caz, limba este obiectul de studiu, iar statistica reprezintă numai instrumentul care permite o investigație mai bună, o caracterizare mai completă a fenomenului lingvistic considerat.

Pe lângă avantajele pe care le oferă un asemenea tip de analiză, ea prezintă și pericole. În acest sens trebuie să considerăm semnificative numai unele elemente (ne gîndim de exemplu, la cele care depășesc un anumit grad de aproximație permis)⁴. De asemenea, rezultatele nu pot fi extrapolate fără măsură (trebuie să ne ferim, de exemplu, să considerăm datele obținute pe baza cercetării unui singur stil drept caracteristice întregii limbi). Altfel exprimat, aceste cerințe reclamă folosirea corectă a metodei.

În primul caz, greșeala poate fi evitată dacă lingvistul are cîteva noțiuni de matematică, pentru a putea înțelege ce e semnificativ și ce nu e.

În al doilea caz, cercetătorul trebuie să aibă în vedere eșantioanele de limbă de la care a pornit și faptul că în ansamblul limbii sînt reprezentative numai unele dintre ele.

³ Cf. și David W. Reed, *A Statistical Approach to Quantitative Linguistic Analysis*, p. 235, 247. O interpretare a celor două aspecte — calitativ-cantitativ — s-a cerut de altfel, încă de la primele discuții asupra metodei statistice: „The importance of quantitative investigations to linguistics is measured by the degree in which it is possible to set up a quantitative system attached to the qualitative system of language”. Cf. intervenția lui Henning Spang-Hanssen, în *Actes du sixième Congrès International des linguistes, Paris, juillet, 1948*, p. 89.

⁴ Este vorba de evenimentele care depășesc limita acceptată pentru erorile standard. Vezi mai departe.

b) Într-o altă variantă în care este concepută lingvistica statistică (și pe care poate ar fi mai bine s-o numim statistică lingvistică) se oferă un loc mai important matematicii. Studiile elaborate într-o asemenea perspectivă privesc limba dintr-un unghi foarte larg, considerînd-o ca pe un fenomen oarecare, neglijînd ce are ea specific. Se poate obiecta, desigur, că nu există aspecte despre care să se fi demonstrat că sînt întîlnite *numai* în limbă. Dar există o „ierarhie” a trăsăturilor de care considerăm că trebuie să se țină seama în orice descriere care urmărește caracterizarea unui fenomen. Acest principiu trebuie respectat și în lingvistică, oricît de „generală” ar fi descrierea. Cercetările la care ne referim urmăresc mai curînd stabilirea unor asemănări între diverse discipline, ca lingvistică, economie, biologie etc., decît clarificarea unor probleme specifice lingvistice. În acest caz, trebuie să facem următoarea precizare: criteriul de apreciere nu e propriu-zis ponderea care se oferă matematicii, ci modul mai mult sau mai puțin natural, corect, în care ea este aplicată.

Pe de altă parte, excesul de matematizare face adeseori inaccesibile rezultatele obținute; facem această precizare, referindu-ne la acele formulări inutile de complicate (sub raport matematic) care nu aduc informații *semnificative* pentru cercetarea limbii. Speculațiile matematice obținute pe această cale nu sînt totdeauna utile pentru lingviști ⁵.

Mai trebuie să reținem o serie de distincții făcute de unii teoreticieni, care au *definit* lingvistica statistică în raport cu alte ramuri ale lingvisticii matematice.

B. Mandelbrot a introdus termenul de „lingvistică statistică macroscopică” (*linguistique statistique macroscopique*). Această disciplină urmează să studieze caracterul global, molar, macroscopic, al unor structuri, care se pretează la explicații în lumina legii numerelor mari, de tipul celor care intervin în teoria informației ⁶.

J. Bar-Hillel face această delimitare: lingvistica *matematică* se ocupă cu aspectele teoretice ale limbii, cu „teoria sistemelor formale”; lingvistica *statistică* lucrează cu numere și statistici (*deals with numbers and statistics*). În cazul cînd se alege denumirea „lingvistică matematică” pentru domeniul mai larg (*genus*), din care lingvistica

⁵ Yvan Lebrun, *Linguistique quantitative*, (în) *Statistique et analyse linguistique*, p. 105—109.

⁶ Benoit Mandelbrot, *Linguistique statistique macroscopique* (în) L. Apostel, B. Mandelbrot, A. Morf, *Logique, langage et théorie de l'information*, p. 1.

statistică este o ramură (*species*), celelalte ramuri ar trebui numite „lingvistică algebrică”⁷.

Pentru H. Karlgren, distanța dintre gândirea statistică și cea algebrică nu este însă atât de mare cum o vede J. Bar-Hillel: ele nu se exclud reciproc⁸.

Din cele câteva precizări de mai sus rezultă că introducerea statisticii în lingvistică conduce la posibilitatea unor caracterizări cantitative și deci și la obținerea unui grad mai mare de exactitate, lucru la care tind toate științele contemporane.



0.3. Analizarea limbii cu mijloace cantitative presupune stabilirea citorva elemente care să explice eficiența aplicării teoriei informației în lingvistică.

Din punctul de vedere al teoriei informației, scopul unui sistem este transmiterea informației. Conținutul mesajului, autenticitatea, importanța lui nu se consideră esențiale. Singura proprietate esențială din acest punct de vedere este structura lui statistică⁹.

Cercetarea aspectului cantitativ sau a structurii statistice a unei informații echivalează cu detectarea într-un șir de simboluri numai a frecvenței și a diferitelor sale combinații.

În termenii teoriei informației, limba este considerată un *cod limitat probabilistic*.

Apropierea limbă — cod se poate face în linii mari, limba cunoscând o caracteristică generală a sistemelor codificate, și anume, utilizarea regulilor combinatorii. Astfel, simbolurile elementare (foneme sau litere) se combină între ele spre a forma unități mai mari (cuvinte) — urmînd regulile matematicii combinatorii. De asemenea, există o stabilitate a frecvențelor și o distribuție probabilistică a elementelor lingvistice. Între limbă și teoria informației există câteva elemente comune. Altfel, fără a exista o stabilitate relativă a simbolurilor lingvistice, nu ar putea funcționa predicția, nu ar fi de „ghicit” mesajul, nu ar fi de înțeles o secvență din care lipsesc segmente, și care rămîne în același timp corectă. Observîndu-se acest

⁷ Jehoshua Bar-Hillel, *Language and Information*, p. 185.

⁸ „It has been insinuated, that statistics could be done without if one had a little more „algebraic imagination”. We see no need *a priori* to exclude probabilistic models from algebraic imagination”. Hans Karlgren, *Statistical Content Analysis*, in *Statistique et analyse linguistique*, p. 112.

⁹ E.V. Paducheva, *Information Theory and the Study of Language*, in *Exact Methods in Linguistic Research*, p. 121.

lucru, diversele elemente (fonemele, de exemplu) au putut fi concepute ca fluctuații ale unor probabilități constante pentru fiecare element considerat din limba respectivă.

Este evident însă că apropierea care se poate face între limbă și cod merge numai pînă la un punct ¹⁰, existînd și diferențe fundamentale. Limba este un caz particular de cod, deci e de la sine înțeles că există proprietăți ale limbii care nu aparțin oricărui cod. În codurile artificiale, de exemplu, restricțiile codului sînt explicite, prestabilite și au un caracter convențional. Ele nu se schimbă decît în urma unui acord explicit între cei care utilizează codul, alcătuiind un sistem închis și fixat. Spre deosebire de acestea, convențiile din limbă apar spontan în cursul comunicării, constituind un sistem deschis și accesibil invențiilor naturale, spontane ¹¹.

Este de reținut caracterul natural al codului lingvistic, ca și evoluția lui specifică, strîns legată de istoria, de evoluția societății care vorbește o limbă sau alta.

De asemenea, în timp ce pentru un cod artificial regula de combinare a simbolurilor este trăsătura caracteristică, în limbă există și alte elemente esențiale. Trebuie să avem în vedere și faptul că limba reprezintă un cod particular, cu restricții proprii, de genul limitării lungimii cuvintelor la un număr oarecare de unități, al preferinței sau repulsiei pentru anumite secvențe de foneme ¹².

De aceea, privind limba numai ca pe un cod, nu putem în nici un caz epuiza problemele pe care le ridică analiza ei. La aprecierile cantitative formale, trebuie să se adauge totdeauna interpretarea lingvistică a faptelor, care transformă simplele observații în date utile pentru teoria limbii.



0.4. Observații istorice asupra lingvisticii matematice în ansamblul raporturilor ei cu structuralismul, ca și a datelor anterioare care au vreo relație cu această disciplină, au fost formulate.

¹⁰ Unii autori constată, de altfel, insuficiența tratării limbii ca un cod: „Summing up this discussion on the ways in which code is conceived with regard to language, it may be said that for the time being the notion of „code” is too complex or too vague to serve as a basis of a uniform study of linguistic matters”. Henning Spang-Hanssen, *Code Theory and „Discrete Mathematics” in Phonology*, p. 53.

¹¹ Pierre Guiraud, *La mécanisation de l'analyse quantitative en linguistique*, p. 37.

¹² G. Herdan, *Language as Choice and Chance*, p. 131.

Întreprinzînd o cercetare de lingvistică statistică, ne propunem să schiţăm cîteva elemente şi despre acest domeniu, în care s-a lucrat, cel puţin în ultimul timp, atît de mult.

Data la care s-a stabilit pentru prima oară o relaţie între statistică şi lingvistică se pierde în timp. O serie de autori¹³ pun pe seama unui asemenea raport însăşi sistematizarea lingvistică actuală. Clasificarea categoriilor şi formelor gramaticale trebuie să se fi efectuat printr-o operaţie de selecţie în materialul vast al limbilor lumii. Desigur, la început n-a fost vorba de aplicarea sistematică a metodei cantitative, ci numai de utilizarea unor consideraţii statistice implicite¹⁴.

După unii autori, statistica lingvistică riguroasă, ştiinţifică, ar începe abia pe la anul 1940, cu lucrările lui Zipf, Yule, Ross. „Pînă atunci exista o lingvistică cantitativă, dar ea nu merită numele de statistică”¹⁵.

După alţi autori, primul nume marcant pentru statistica ştiinţifică ar fi Markov — care a formulat, în 1913, legile (numerice) de succesiune ale fonemelor în secvenţele pe care azi le numim „lanţuri Markov”¹⁶.

În orice caz, după publicarea *Cursului de lingvistică generală* al lui F. de Saussure (1916), cînd ştiinţa limbii a intrat într-o nouă etapă, şcolile lingvistice structuraliste au inclus în preocupările lor şi probleme de lingvistică cantitativă.

Dintre membrii *şcolii de la Praga*, Vilém Mathesius este primul care atrage atenţia asupra necesităţii de a aborda limba din punct de vedere cantitativ. În articolele sale *La structure phonologique du*

¹³ G. Herdan, *Language as Choice and Chance*, p. 1.

¹⁴ Pierre Guiraud, în *Bibliographie critique de la statistique linguistique*, urmăreşte îndeaproape istoricul acestei discipline. El notează atît încercările mai vechi, ale criptografilor din sec. al XV-lea, de a detecta stabilitatea unor elemente lingvistice (litere) în jurul unor valori, cît şi lucrările mult mai noi, ale stenografilor, care au atras atenţia lingviştilor asupra frecvenţei sunetelor. Desigur, au mai existat şi alte cercetări, anterioare aceluia moment de neobişnuită amploare a studiilor cantitative care este pus în legătură cu dezvoltarea cercetărilor din teoria informaţiei. Pentru unele observaţii istorice, vezi şi L. Kaiser, *A Statistical Research of Spoken Language (Dutch)*, p. 235—248.

¹⁵ Cf. P. Guiraud, *Problèmes et méthodes de la statistique linguistique*, p. 6 şi L. Kaiser, *op. cit.*, p. 246. Nu este mai puţin adevărat că legile lui Zipf (şi ale lui Yule) au suscitat numeroase discuţii ulterioare. Pentru unele comentarii, cf. Benoit Mandelbrot, *op. cit.*, p. 22 ş.u.; Warren Plath, *Mathematical Linguistics*, p. 24—26. La Colocviul de statistică lingvistică de la Strasbourg, G. Herdan a calificat legea lui Zipf ca „învechită în zilele noastre, nefiind o lege în sens teoretic şi fiind empiric neadevărată”. Cf. *Statistique et analyse linguistique*, p. 54.

¹⁶ Milka Ivič, *Trends in Linguistics*, p. 216.

lexique du tchèque moderne, TCLP I, 1929, *Zum Probleme der Belastungs- und Kombinationsfähigkeit der Phoneme*, TCLP, 1931¹⁷, el distinge trei aspecte ale utilizării calității fonice în limbă, și anume: în sistem, în structura cuvintelor și a grupurilor de cuvinte, în vorbirea concretă.

Calitatea fonemelor, a combinațiilor lor în unități mai mari, cuvintele, nu este suficientă, după V. Mathesius, pentru a caracteriza o limbă. Este cazul să se adauge considerații cantitative, spune el, indicându-se procente de utilizare ale acestor tipuri, procentaje care se calculează statistic¹⁸. El însuși a întreprins un asemenea studiu pentru limbile cehă și germană. Ulterior, în urma sugestiilor sale, B. Trnka (1935) a analizat statistic cuvintele engleze de 1 și 2 silabe, iar J. Vachek (1940), pe cele din limba cehă — de 5 foneme¹⁹.

N. S. Troubetzkoy, în *Principes de phonologie* (1939) consacră un capitol întreg — „De la statistique phonologique” — statisticii fonologice. El consemnează realizările anterioare (lucrările lui V. Mathesius, Trnka, Twadell, problemele generale enunțate de G. K. Zipf). În plus, notînd că, pînă la scrierea cărții sale, nu s-a conturat încă o metodă statistică unică, schițează cîteva idei.

După Troubetzkoy, semnificația statisticii în fonologie este următoarea :

a. să arate care este frecvența unor elemente fonologice determinate (foneme, grupuri de foneme, tipuri de morfeme sau cuvinte) într-o limbă dată, în texte ; b. să stabilească randamentul funcțional al acestor elemente, sau al unor opoziții fonologice determinate, în dicționare²⁰. El observă că cifrele²¹ absolute au mai puțină valoare :

¹⁷ Ambele articole sînt reeditate în : Joseph Vachek, *A Prague School Reader*....

¹⁸ « Mais le caractère phonologique d'une langue n'est pas déterminé seulement par la quantité des éléments phonologiques et le volume de leurs possibilités d'utilisation, mais aussi par la proportion dans laquelle il est fait usage de ces possibilités ». Vilem Mathesius, *La structure phonologique du lexique...* (în) J. Vachek, *A Prague School Reader*..., p. 175. Cf. și : „Zur phonologischen Charakteristik einer Sprache genügt es nicht, ihren Vorrat von Phonemen und phonologischen Merkmalen festzustellen : man muss auch die Intensität untersuchen, mit der die einzelnen phonologischen Einheiten in der behandelten Sprache verwandt werden. So behauptet in der Phonologie neben der Qualitäten auch die quantitative Analyse ihren Platz”. V. Mathesius, *Zum Belastungs- und Kombinationsfähigkeit...* J. Vachek, *A Prague School Reader*..., 177.

¹⁹ Vezi pentru discuțiile despre lingvistică statistică în școala pragheză — Joseph Vachek, *The Linguistic School of Prague*, p. 64—68.

²⁰ N.S. Troubetzkoy, *Principes de phonologie*, p. 276—279.

²¹ « Seul le rapport entre ce chiffre et le chiffre de fréquence attendu théoriquement possède une valeur véritable ». N.S. Troubetzkoy, *op. cit.*, p. 284.



importantă este proporția dintre datele cuantificate, astfel încât trebuie exprimate probabilități, ținând seama de toate regulile de combinare și de neutralizare.

Statistica fonologică trebuie să acorde importanță nu numai frecvenței de apariție a fonemelor în general, ci și secvenței lor în poziții determinate.

Troubetzkoy mai discută dificultățile alegerii unor eșantioane reprezentative pentru limba pe care o studiem, alegere care trebuie făcută paralel cu stabilirea a ce este și ce nu este element stilistic.

El se referă, de asemenea, la principiile de statistică lingvistică ale lui J. Van Ginneken (asemănarea lingvistică dintre vorbitorii aceleiași limbi se datorează unor elemente comune, rasiale) și ale lui Zipf (teoria minimului efort, potrivit căreia fenomenele cele mai puțin complicate sînt cele mai frecvente). Troubetzkoy pune în evidență caracterul biologic al acestor teorii și formularea lor pe baze extra-lingvistice, observînd totuși că legea lui Zipf poate fi înțeleasă, pînă la un punct, și pe plan lingvistic.

Leonard Bloomfield, principalul reprezentant al *școlii descriptive americane* (școala de la Yale), formulează și el cîteva observații cantitative asupra structurii limbilor. Descriînd structura fonetică a limbilor, în volumul *Language* (1933), Bloomfield notează deosebiri între frecvențele fonemelor din diverse limbi, alături de legi comune mai multor limbi (de exemplu, frecvența consoanelor surde superioară frecvenței consoanelor sonore). Se mai remarcă existența unui raport între frecvența mare a unor foneme și frecvența cuvintelor care le conțin (de exemplu, fonemul /ð/ din cuvîntul englez *the*). Asemenea observații, ca și constatarea că mai ales nespecialiștii s-au lansat în cercetări statistice, îl fac pe lingvistul american să afirme necesitatea unui studiu serios în acest domeniu ²².

În principiile fundamentale ale *școlii glosematice* nu figurează discuții despre lingvistica statistică. Întîlnim însă numele unui membru reprezentativ al școlii de la Copenhaga, chiar în Actele Congresului al VI-lea de lingvistică, congres la care s-a constituit Comitetul pentru lingvistică cantitativă. Henning Spang-Hanssen, la care ne referim, a prezentat la acest congres raportul *Remarks Concerning the Importance and Methodology of Quantitative Investigation* ²³. În

²² „A serious study of this matter is much to be desired”. Leonard Bloomfield, *Language*, p. 137.

²³ În *Actes du sixième Congrès International des Linguistes, Paris, juillet, 1948*, p. 370–390.

expunerea sa, Spang-Hanssen pornește de la cele două laturi ale limbii, expresia și conținutul, observînd că statistica poate stabili anumite afinități obiective între clasele de elemente. Aportul studiilor statistice în lingvistică va fi apreciat în măsura în care este posibil să se contureze un sistem cantitativ, alături de sistemul calitativ al limbii²⁴. Din aceste considerații sumare, se observă că, de fapt, toate școlile structuraliste au o atitudine favorabilă față de lingvistica statistică²⁵. La al VI-lea congres internațional de lingvistică, Paris, 1948, s-a luat inițiativa creării unui „comitet pentru lingvistica cantitativă”, avînd următoarele sarcini: **a.** să stabilească principii concrete empirice pentru cercetarea cantitativă a sistemelor lingvistice; **b.** să descopere caracteristici cantitative ale limbilor în vederea comparării lor; **c.** să releve principiile specifice care guvernează (determină) distribuția frecvențelor la elementele lingvistice; **d.** să stabilească relațiile dintre distribuția acestor frecvențe și aspectele calitative ale vorbirii individuale sau colective²⁶.

În 1951, în S.U.A. s-a ținut primul curs de lingvistică statistică.

Așa se detașează lingvistica cantitativă ca o ramură oarecum independentă în ansamblul lingvisticii matematice.

În prezent se organizează manifestări științifice internaționale consacrate exclusiv acestei discipline. În noiembrie 1960 a avut loc la Sundbyholm Manor simpozionul scandinav de statistică lingvistică. La această întrunire s-au discutat probleme diverse: relația dintre rangul fonemelor și distribuție, probabilitatea și clasificarea structurală, combinabilitatea fonemelor etc.²⁷.

La o altă întrunire, Colocviul de statistică și analiză lingvistică de la Strasbourg, raportorii Charles Müller și B. Pottier, au caracterizat cu bună dreptate etapa actuală astfel: colocviul ilustrează un

²⁴ Interesul pentru rezolvarea unor probleme lingvistice teoretice cu ajutorul calculului probabilităților și al clasificării structurale se reflectă, de altfel, și în unele dintre studiile sale ulterioare. De exemplu: „In general it seems plausible that the highly refined quantitative model developed by information theory, if applied with due consideration of linguistic interrelationships, will form a genuine supplement to the principles of linguistic analysis in terms of distribution”. Henning Spang-Hanssen, *Typological and Statistical Aspects of Distribution as a Criterion in Linguistic Analysis*, p. 194.

²⁵ Cf. și Jiří Krámský, *Quantitative Phonemics in the Last Decade*, p. 166.

²⁶ Cf. *Actes de sixième Congrès international des linguistes, Paris, juillet, 1948*; John B. Carroll, *The Study of Language*, p. 61—64.

²⁷ Programul integral al simpozionului este reprodus în *Statistical Methods in Linguistics*, Språgförlaget Skriptor, Stockholm, Sweden, 1, 1962, p. 10—11. Revista *Statistical Methods in Linguistics* reprezintă actualmente o publicație permanentă de statistică lingvistică.

moment de echilibru survenit după perioada de exces. Lingviștii știu acum că nu pot neglija aspectul cantitativ al limbii : ei se întreabă numai în ce măsură trebuie considerat acest aspect ca pertinent, de la caz la caz, în funcție de cîmpurile specifice de cercetare ²⁸.

0.5.1. S-a discutat mult pentru și împotriva tehnicilor probabile, lingvistica statistică avînd, ca și alte discipline, entuziaști și adversari.

Un argument puternic în favoarea lingvisticii cantitative este însuși faptul că ea reprezintă singura metodă aplicabilă atunci cînd avem de studiat ansambluri mari de elemente ²⁹. Or, lingvistica lucrează cu fapte multe și care se produc simultan, domeniul ei de cercetare este vast. De aici, entuziasmul unor cercetători pentru lingvistica statistică devine total ³⁰.

Cazul extrem este cel în care se ajunge pînă la a pune problema în termeni de *sau ... sau* : sau lingvistică statistică, sau gramatici generative ³¹.

Pentru a împărtăși acest punct de vedere, ar însemna să cerem lingvisticii cantitative să înlocuiască toate cercetările de orice altă natură și să rezolve indiferent ce probleme.

0.5.2. Adversarii lingvisticii statistice au formulat o serie de obiecții, care se pot rezuma astfel :

I. Obiecțiile teoretice : natura esențial calitativă a limbii, complexitatea ei și caracterul independent față de determinismul statistic.

Or, aceste obiecții neglijează faptul că același fenomen poate avea un dublu caracter : calitativ și cantitativ. Semnele lingvistice

²⁸ Vezi *Statistique et analyse linguistique*, p. 1.

²⁹ Cf. și Yvan Lebrun, *Linguistique quantitative*, în *Statistique et analyse linguistique*, p. 107 ; Pierre Guiraud, *Problèmes et méthodes de la statistique linguistique*, p. 1—2. Acesta este, de altfel și punctul de vedere al lui G. Herdan, în toate lucrările sale (vezi bibliografia).

³⁰ «La linguistique est la science statistique type : les statisticiens le savent bien : la plus part des linguistes l'ignorent encore» — afirmă Pierre Guiraud (*Problèmes et méthodes de la statistique linguistique*, p. 15).

³¹ „If we wish to continue in the tradition of Saussure and Bloomfield, the direction in which linguistics as a science must develop is not towards a generative grammar, but towards statistical linguistics, or, more specifically, not towards a mere re-formulation of already known linguistic facts in a way that would only satisfy the logician, but in the direction of linguistics as a branch of applied mathematics, which, as has been abundantly shown in recent years, yields new results in the form of hitherto unknown relationships. It is not re-definition of known relations that enriches scientific knowledge, but the discovery of new methods of investigation”. G. Herdan, *Quantitative Linguistics or Generative Grammar*, p. 65.

sint creații individuale, în principiu, dar și colective, întrucit valoarea lor stă tocmai în măsura în care se lasă primite, transmise etc. Statistica are rolul de a interpreta la nivel colectiv elemente ce nu se pot nici repera, nici defini la nivel individual.

În ce privește complexitatea limbii, se poate răspunde că însuși scopul metodei statistice este analiza faptelor complexe ale căror componente nu pot fi imediat descifrate prin observare. Faptele de observație arată că, deși expresiile lingvistice pot apărea individual ca produse ale unor alegeri libere, în înțelesul obișnuit al cuvîntului, în general semnele se distribuie după scheme aleatorii. Statistica dezvăluie, deci, singularitatea faptelor izolate și determinismul ansamblului, al limbii ca un organism total.

II. Obiecțiile practice: analiza numerică cere numărători și calcule enorme.

Dar o bună parte din munca laborioasă pe care o solicită studiile cantitative se poate efectua în mod automat, cu mașini special construite sau adaptate ³².

Este de observat că atât obiecțiile teoretice, cît și cele practice, se bazează pe eventualele dificultăți de cercetare, de o natură sau alta, care pot surveni în aplicarea metodei statistice.

Nu de la eforturi (ipotetice sau reale) trebuie să pornim însă atunci cînd discutăm validitatea unei metode. Important este dacă ea e *adekvată* obiectului de cercetare, dacă permite obținerea unor rezultate mai exacte (în cazul nostru, mai formalizate, formalizarea incluzînd cuantificarea propriu-zisă) sau degajarea unor aspecte noi într-un domeniu științific.

0.5.3. O serie de discuții s-au referit *nu* la descrierile cantitative generale ale unui fenomen sau altuia, ci la cazul particular a *modelelor* lingvistice statistice. Astfel, Noam Chomsky a arătat că modelul statistic, considerațiile probabilistice nu contribuie la înțelegerea valorii modelului gramatical; frecvența întrebuițării unor construcții nu le conferă un grad mai mare sau mai mic de gramaticalitate ³³. Modelul statistic ar deveni inadecvat pentru acest do-

³² Pentru formularea principalelor obiecții aduse metodei statistice în lingvistică ca și pentru o serie de discuții în acest subiect, vezi Pierre Guiraud, *Problèmes et méthodes de la statistique linguistique*, p. 19—23.

³³ Noam Chomsky, *Syntactic Structures*, p. 16—17. Pe de altă parte, V.M. Zolotarev, a construit un model probabilistic — privind structurile sintactice, pe care le analizează ca elemente într-un lanț Markov. Cf. N.D. Andreev, *Models as a Tool in the Development of Linguistic Theory*, p. 189.

menu lingvistic. Și E. V. Paducheva observă că date fiind particularitățile matematice ale teoriei informației, ea nu oferă metode de studiu decât pentru o parte din aspectele limbii, și anume pentru cele formale, sau, mai exact, codificabile. Este cazul fonemelor și, la alt nivel, al literelor care se succedă unele după altele, în secvență liniară ³⁴. I. I. Revzin demonstrează strâns legătură dintre modelare și statistică. El motivează necesitatea introducerii criteriilor statistice în crearea modelelor lingvistice astfel :

a. Modelul are în vedere un obiect ideal; corespondența dintre acest obiect ideal și relațiile existente este totdeauna mai mult sau mai puțin relativă, iar gradul de corespondență trebuie definit statistic.

b. Orice model utilizează implicit sau explicit un material lingvistic inițial, care cuprinde o parte din limbă. Chiar atunci când modelul se referă la o mulțime infinită de propoziții, în trecerea de la model la analiza materialului concret, reîntâlnim o anumită mulțime finită, sau, din punct de vedere statistic, un anumit eșantion, o probă. Posibilitatea de a afla care este mărimea acestui eșantion este o problemă tipic statistică.

c. Modelul operează cu termeni logici de tipul *toți* (de exemplu, „toate frazele de tipul...”). De îndată ce se trece la aplicarea modelului, corelativul acestui termen devine termenul *un număr suficient de*.

d. Când avem de-a face cu mai multe modele care descriu același obiect, urmează să stabilim care dintre ele este mai eficient. Eficiența se apreciază atât din punct de vedere statistic, cât și din punct de vedere logic.

e. Interpretarea modelului: conceptul fundamental al modelării „înregistrarea unei fraze” trebuie de asemenea interpretat pe baza unor considerații statistice.

Asemenea fapte îl fac pe Revzin să remarce că există o legătură esențială între modelare și statistică ³⁵.

După R. Moreau, datorită, pe de o parte, faptului că pe baza criteriilor statistice ne putem referi la un material lingvistic vast, iar pe de altă parte, pentru că acest material nu este fără limite, va trebui să lucrăm cu un model care îndeplinește mai multe condiții.

³⁴ E.V. Paducheva, *Information Theory and the Study of Language*, în *Exact Methods in Linguistic Research*, p. 175.

³⁵ „It ought always be remembered that there is an essential connection between making models and Statistics”, I.I. Revzin, *Models of language*, p. 11–12.

Astfel, avem posibilitatea de a construi, prin sondaj, un model redus, dintr-un ansamblu care are prea multe elemente pentru a putea fi studiat în întregime. Pornind de la acest model redus, urmează să apreciem unele caracteristici ale întregului ansamblu (medie, abatere - tip etc.). Cu același model redus, putem verifica ipoteze făcute asupra întregului ansamblu ³⁶.

Obiecțiile expuse de noi mai sus se referă la metodă în ansamblul ei.

Trebuie să notăm, de asemenea, că metoda statistică a fost discreditată în ultimul timp, mai ales pentru felul în care s-a aplicat și pentru modul în care s-au *interpretat* rezultatele ei. Cele mai multe discuții se referă, de aceea, la o lucrare sau alta. Lingviștii, în general, au acceptat însă *posibilitatea* ca cercetările de acest tip să ofere date relevante ³⁷.

0.5.4. Ca și alte ramuri ale lingvisticii moderne, lingvistica statistică — în calitate de disciplină constituită — se reclamă ca având bazele teoretice în concepția despre limbă a lui F. de Saussure. Fără a afirma că Saussure avea o concepție „cantitativă” asupra limbii, lingviștii statisticieni selectează din lingvistica descriptivă câteva principii ca : a) raportul de independență dintre sunetele și sensul unui cuvânt (raportul signifiant/signifié) ; b) cele două aspecte ale limbajului (*langage*) dintre care esențială este limba (*langue*), iar vorbirea (*parole*) este elementul secundar. Raportul dintre acești factori se mai poate enunța și : „limba reprezintă limbajul minus vorbirea” ³⁸.

Obiectul cercetării statistice — ca și al întregii lingvistici propriu-zise ³⁹ — îl constituie limba în ansamblul ei.

³⁶ Cf. *Intervention de R. Moreau, in Statistique et analyse linguistique*, p. 125.

³⁷ Notăm, printre altele : „Although statistical properties of language and explanatory models for observed uniformities are certainly worth studying, it seems that such investigations have not yet reached the point where they make a significant contribution to the understanding of linguistic process”. Noam Chomsky, *recenzia* la Vitold Belevitch, *Langage des machines et langage humain*, p. 105 ; „This above is not to be construed a blanket condemnation of statistical studies of language. I do not rule out, by no means, the possibility that statistical hypotheses of deep significance for linguistics may yet be discovered”. Morris Halle, *recenzia* la G. Herdan, *Language as Choice and Chance*, p. 23. Și M.A.K. Halliday recunoaște importanța cercetării statistice, fără a putea accepta însă părerea lui Herdan că, numai aceasta va transforma lingvistica structurală într-o știință. „None of these [articles] negates the important role of statistics in linguistic description”. M.A.K. Halliday, *recenzia* la G. Herdan, *Language as Choice and Chance*, p. 149.

³⁸ Ferdinand de Saussure, *Cours de linguistique générale*, p. 122.

³⁹ Id. *ib.*, p. 38.

Calculul probabilităților indică în plus modul în care utilizează o limbă materialul din care este alcătuită, proporția dintre elementele componente.

Înțelegînd limba ca un sistem — (în termeni saussurieni, « un système ou tout se tient »), lingvistica statistică are în vedere descrierea limbii ca un sistem de solidaritate între formele aparținînd unui anumit *nivel*. Indicii statistici ai descrierii de acest tip caracterizează de fapt structura fiecărui nivel lingvistic privit ca un întreg.



0.6.1. Nu mai este necesar să amintim că metoda cantitativă se aplică într-o serie de sectoare lingvistice. Precizăm că expunerea noastră se oprește doar la acele principii generale, definiții etc., care privesc direct sau indirect sistemul *fonologic*. Nu ne propunem discutarea tuturor raporturilor statistică/lingvistică, în planuri ca vocabularul, sintaxa etc.⁴⁰. Am intitulat lucrarea noastră „*fonostatistică*”, dînd termenului accepțiunea „relativ la frecvența fonemelor”⁴¹.

De asemenea, am considerat termenul *statistic* sinonim cu cantitativ⁴². O delimitare riguroasă a domeniilor ne-ar conduce la distincții între aspectele statistice, aspectele numerice, aspectele cantitative. Nu orice descriere cantitativă este o descriere statistică (spre a da numai un exemplu, aritmetica, deși este o disciplină cantitativă, nu aparține statisticii). În cazul de față, pentru a nu complica lucrurile, cît și pentru că distincțiile sînt totuși facultative, am preferat echivalența termenilor statistic — cantitativ. În același timp, am urmărit numai problemele care aparțin domeniului statistic, adică faptele cu caracter aleatoriu.

În general, tipologia fonologică cantitativă poate avea ca obiect fie fonemele, considerate ca unități indivizibile, fie relațiile dintre trăsăturile distinctive ale fonemelor, concepute ca struc-

⁴⁰ S-a observat, de altfel, și cu alte ocazii, că din lingvistică, apreciată drept cea mai exactă știință umanistă, fonologia reprezintă sectorul cel mai exact. Aici, precizia formulărilor a putut fi comparată cu cea din matematică. M.V. Mačavariani, *On the Relationship between Mathematics and Linguistics*, p. 27.

⁴¹ «Phonostatistique signifie relatif à la fréquence des phonèmes». P. Guiraud, *Problèmes et méthodes de la statistique linguistique*, p. 99.

⁴² „All linguistic research which rely on the statistical method are called statistical or quantitative; even linguistics in general is qualified as *statistical* or *quantitative* in so far as language investigators make use of statistics”. Milka Ivič, *op. cit.*, p. 220—221.

turi⁴³. Cercetarea noastră are în vedere relațiile dintre unitățile segmentate, fonemele, celălalt aspect (trăsăturile dinstinctive) urmînd să fie luat în considerație după stabilirea cîtorva concluzii privitoare la entitățile fonematice, concepute ca foneme sau ca grupuri de foneme. Pe de altă parte, se știe că orice unitate minimală, degajată pe baza segmentării și a analizei fonematice, capătă o valoare lingvistică prin relațiile în care este angrenată în mod necesar. Un aspect îl constituie raporturile din planul paradigmatic. Cantitatea de informație pe care o poartă fiecare unitate minimală este cu atît mai mare cu cît este sistemul mai bogat. În același timp, ca elemente ale unui lanț de unități de expresie, într-o sintagmă, unitățile minimale intră în relații reciproce, sintagmatice. Structurile sintagmatice sînt importante în stabilirea valorii informative a fonemelor⁴⁴. Regulile de distribuție modifică de obicei cantitatea de informație pe fonem.

În toate limbile, posibilitățile de construire a unor unități mai mari din entități primare sînt limitate. Cum noi nu vom discuta decît problemele legate de *înlănțuirea* fonemelor (cu excepția capitoului în care ilustrăm frecvența *fiecărui* fonem românesc, pentru a avea o imagine de ansamblu), cercetarea noastră se menține în planul relațiilor sintagmatice. O descriere paradigmatică — pe care nu intenționăm s-o efectuăm deocamdată — ar pune în lumină și alte aspecte⁴⁵.

Am selectat elementele fonologice pe care le-am considerat mai caracteristice pentru obținerea unei imagini de ansamblu⁴⁶.

⁴³ Roman Jakobson (with Collin Cherry and Morris Halle), *Toward the Logical Description of Languages in their Phonemic Aspect*, I, p. 457. Pentru discutarea unor aspecte, vezi și Andrei Avram, *Asupra tipologiei fonologice cantitative*, p. 323—326. De asemenea, pentru tipurile de modele fonologice vezi Solomon Marcus, *Le modelage mathématique en phonologie*, p. 109—116.

⁴⁴ „Even the structure of the syntagm, and, more generally, of the complex expression units (syllables, groups, etc.) are of importance for the phonemes“. B. Malmberg, *Structural Linguistics and Human Communication*, p. 128.

⁴⁵ Pentru prezentarea celor mai importante puncte de vedere în această problemă ca și pentru propunerea unor noi criterii de apreciere, vezi Andrei Avram, *Some Thoughts on the Functional Yield of Phonemic Oppositions*, p. 40—47.

⁴⁶ Unele dintre acestea coincid cu elementele indicate de P. Menzerath și W. Meyer-Eppler ca relevante într-o cercetare tipologică. Obiectivele propuse de ei sînt: a) numărul de silabe în cuvînt — care permite obținerea tipurilor de cuvinte (*Worttypen*); b) numărul fonemelor în cuvînt — permite obținerea claselor de cuvinte (*Wortklassen*); c) ordinea fonemelor în cuvînt — duce la obținerea tipurilor de forme (*Formtypen*); d) numărul și ordinea fonemelor în cadrul silabei (*Sprachtypologische Untersuchungen*, p. 57—58).

Atenția noastră s-a oprit fie la *raportul cantitativ dintre unități* : vocale/consoane, grupuri de consoane / grupuri de vocale (și în interiorul categoriilor, subdiviziuni ca : grupuri de două consoane / grupuri de trei sau mai multe consoane etc.) și de *relația de tranziție* (între vocale, între consoane, între vocale și consoane, între consoane și vocale) ; fie la *poziția* segmentelor fonematice (inițială, medială, finală) : fie, în sfârșit, la *dimensiunile* elementelor luate în discuție (*câte* unități poate cuprinde un grup fonematic omogen ; care este lungimea celor mai frecvente tipuri silabice ; lungimea în silabe — a cuvintelor). De asemenea, în unele cazuri am calculat *entropia* obținută în diverse condiții. Atunci cînd a fost posibil, am efectuat comparații cu alte limbi.

Ne-am propus astfel să cercetăm o serie de elemente legate între ele, semnificative pentru nivelul fonologic, și să obținem o serie de informații care să nu constituie simple „fapte izolate de limbă”⁴⁷.

0.6.2. În lucrarea noastră ne-am bazat pe materialul oferit de texte românești scrise. Procedînd astfel, nu considerăm nicidecum că am epuizat sfera de cercetare în domeniul limbii române. Pentru un studiu complet urmează să obținem și date caracteristice limbii vorbite, ceea ce va constitui obiectul unei lucrări viitoare.

După unii autori însă, este chiar recomandabil ca investigațiile de această natură să înceapă cu limba scrisă deoarece cuvintele scrise ilustrează forme stabile, cunoscute și pe care limba le păstrează mai mult timp.

În alt plan, ar mai fi de adăugat că studierea unităților lingvistice scrise este esențială pentru stabilirea unei relații proprii (*proper understanding*) în comunicarea dintre mașină și om și dintre om și mașină⁴⁸.



0.6.3. Înainte de a începe descrierea cantitativă a unor elemente fonologice, este necesar să facem cîteva precizări privitoare

Dintre aspectele care caracterizează nivelul fonologic, G. Herdan reține ca aspecte importante : distribuția statistică a fonemelor în limbă ; entropia unui sistem fonemic, în sensul teoriei informației (*Statistics of Phonemic Systems*, p. 435—439).

⁴⁷ “If statistical or other mathematical tools are to be employed in linguistics it is essential to show that these tools can handle problems of significance for linguistics, and not just isolated facts of language”, Morris Halle, *recenzia* la G. Herdan, *Language as Choice and Chance*, p. 22.

⁴⁸ Cf. James L. Dolby și Howard L. Resnikoff, *On the Structure of Written English Words*, p. 167.

la noțiunile cu care vom lucra. Vom încerca să definim o serie de termeni.

Frecvența de apariție a unui fonem se definește ca proporția dintre numărul de apariții efective ale fonemului respectiv și numărul total de apariții teoretic posibile⁴⁹. În statistica noastră, lucrăm cu frecvențe relative. *Frecvența relativă* a unui eveniment A este raportul dintre numărul probelor v în care evenimentul A s-a produs și numărul total de n probe diferite⁵⁰.

Probabilitatea unui element reprezintă raportul dintre numărul de apariții ale elementului într-un text și numărul total al elementelor textului⁵¹. O altă definiție a probabilității este: probabilitatea unui eveniment este egală cu raportul dintre numărul de rezultate echiprobabile favorabile evenimentului considerat și numărul total de rezultate echiprobabile⁵².

În general, frecvența și probabilitatea unui eveniment nu trebuie confundate. Observații îndelungate au arătat că, dacă se repetă un experiment în aceleași condiții de un număr suficient de mare de ori, atunci frecvența relativă este stabilă, adică oscilează în jurul probabilității⁵³.

Există însă modele în care frecvența nu converge spre probabilitate.

Prin probabilitate se caracterizează diverse evenimente, fiecare în parte, independent de frecvența cu care se poate eventual asocia probabilitatea.

Pentru ca frecvența să aproximeze cât mai bine probabilitatea, este necesar să fie satisfăcute următoarele condiții:

a) experimentele care au ca rezultat producerea evenimentelor statistice să fie efectuate independent unele de altele;

⁴⁹ Cf. R.M. Frumkina, *The Application of Statistical Methods in Linguistic Research in Exact Methods in Linguistic Research*, p. 87. Notăm și această observație judicioasă asupra frecvenței: « Le concept de fréquence est utilisé en linguistique sous son double aspect subjectif et quantitatif: subjectif, il permet d'opposer un événement fréquent à un événement rare sans borner inférieurement, à l'aide d'une critère rigoureux, la classe des événements fréquents; quantitatif, il permet de mesurer le nombre d'apparitions d'un événement donné ». R. Moreau, *Au sujet de l'utilisation de la notion de fréquence en linguistique*, p. 140.

⁵⁰ Marius Iosifescu, Gheorghe Mihoc, Radu Theodorescu, *Teoria probabilităților și statistica matematică*, p. 9.

⁵¹ I.M. Iaglom, R.L. Dobrușin, A.M. Iaglom, *Teoria informației și lingvistica*, p. 28.

⁵² A.M. Iaglom, I.M. Iaglom, *Probabilitate și informație*, p. 17.

⁵³ Marius Iosifescu, Gheorghe Mihoc, Radu Theodorescu, *op. cit.*, p. 9.

b) rezultatele să se stabilească pe baza unui număr suficient de experiențe.

De altfel, legitățile proprii fenomenelor de masă se manifestă, în mod firesc, cu atât mai pregnant, cu cât volumul materialului statistic este mai mare. Aceasta se leagă de faptul că orice legitate statistică se manifestă simultan cu acțiunea unei alte legități obiective, și anume, cu legea numerelor mari ⁵⁴.

În lingvistică, elementele cu care lucrăm sînt în general interdependente ⁵⁵. Frecvența unui fonem, de exemplu, este condiționată de diverse contexte.

La o anumită distanță între foneme însă, dependențele sînt diminuate, apoi se anulează — fapt care reflectă un proces ergodic.

Teoria informației măsoară influența la distanță a unităților lingvistice.

În ceea ce privește literele, de exemplu, s-a stabilit că apariția unei litere influențează, în general, aparițiile următoarelor 30 de litere, dar, de la a 31-a literă, influența dispăre. Altfel spus, apariția unei litere este influențată (în medie) de cele 30 de litere imediat precedente ⁵⁶.

Într-un asemenea caz, probabilitatea converge către frecvență, într-o limbă „convențională”, pentru că e vorba de o probabilitate limită.

Precizăm, deci, că atunci cînd vom întrebuița noțiunea de probabilitate, vom subînțelege probabilitatea limită.

Din această cauză nu mai insistăm asupra proprietăților probabilităților ⁵⁷.

După cum am mai arătat, practic, noi lucrăm de fapt cu frecvențe relative și, ca terminologie, nu facem distincția frecvență/probabilitate.

⁵⁴ I. Marinescu, C. Moineagu, R. Niculescu, N. Rancu, V. Urseanu, *Elemente de statistică matematică și aplicațiile ei*, p. 10—11.

⁵⁵ Statistica matematică semnalează existența unor cazuri în care chiar natura fenomenului nu permite realizarea unei selecții independente. Cf. O. Onicescu, Gh. Mihoc, *Lecții de statistică matematică*, p. 578.

⁵⁶ N.G. Burton, J.C. Licklider, *Longrange Constraints in the Statistical Structure of Printed English*. I.M. Iaglom și A.M. Iaglom (*Probabilitate și informație*) extind ipotetic valoarea și pentru alte limbi.

⁵⁷ Pentru înțelegerea elementară a teoriei probabilităților și a unora dintre căile aplicării ei, cf. și B.V. Gnedenko și A.I. Hincin, *Introducere elementară în calculul probabilităților*.

Este de reținut că „noțiunea de probabilitate a unui eveniment nu poate fi despărțită de noțiunea de frecvență relativă”⁵⁸; ele nu reprezintă însă elemente identice. În cele ce urmează, frecvențele au fost exprimate în *procente* (%).

Eroarea standard este un coeficient care măsoară validitatea rezultatelor obținute. Se știe că rezultatele sînt cu atît mai plauzibile cu cît sînt mai largi textele din care s-au extras exemplele. Pentru verificare, au fost propuse o serie de formule dintre care am reținut :

$$\text{Eroare Standard} = \sqrt{\frac{pq}{N}}$$

în care : p = frecvența elementului considerat ; q = frecvența celorlalte elemente din text ; N = numărul tuturor elementelor din text⁵⁹.

În general, se consideră că rezultatele obținute sînt exacte atunci cînd eroarea standard nu depășește cifra 5% (= 0,05).



0.7.1. Pentru excerptarea materialului fonologic, am lucrat cu un număr mare de texte diferite. Toți autorii subliniază importanța varietății materialului și a întinderii lui, pentru a se putea obține date cantitative relevante⁶⁰. Altfel riscăm să rămînem în limitele unor exemple restrînse, fiecare serie de rezultate caracterizînd doar textele pe baza cărora a fost stabilită. Numai la un număr mare de exemple se poate vorbi de „legea numerelor mari”.

Diversitatea textelor este asigurată prin extragerea de eșantioane din mai multe texte. Se aplică astfel metoda selectivă. Din statistica matematică ne însușim principiul de-a opera selecțiile la întîmplare, înțelegînd prin *selecții* grupul de elemente dintr-o colectivitate observat prin probe repetate. Numărul elementelor din grup se numește volumul selecției. Selecției i se mai spune și *eșantion*⁶¹.

În general, am lucrat pe texte de beletristică, din literatura (proza) contemporană ; pentru a avea și elemente de limbă curentă,

⁵⁸ Silviu Guiașu, Radu Theodorescu, *Matematica și informația*, p. 30.

⁵⁹ David W. Reed, *A Statistical Approach to Quantitative Linguistic Analysis*, p. 239.

⁶⁰ Cf. R.M. Frumkina, *The Application of Statistical Methods in Linguistic Research*, în *Exact Methods in Linguistic Research*, p. 88 ; Edwin B. Newman, *Statistical Methods in Phonetics*, p. 123 ; Henning Spang-Hanssen, *Probability and Structural Classification*, p. 69.

⁶¹ H.M. Ionescu, *Statistica matematică*, p. 15.

mai puțin „stilizate”, am adăugat texte din presă (ziare, reviste). Numai în anumite cazuri (pe care le-am specificat), am recurs la lucrări științifice, care se caracterizează printr-un stil aparte.

Textele sînt următoarele :

Pentru stilul beletristic am excerptat material din următoarele volume (fragmente) :

Al. Brătescu-Voinești, *Întuneric și lumină*, Iași, 1914, Liviu Rebreanu, *Nuvele*, Ed. tineretului, 1957, Camil Petrescu, *Patul lui Procust*, Ed. Fundațiilor, 1946, *Caragiale în vremea lui*, ESPLA, 1957, Cezar Petrescu, *Omul din vis*, București, Hortensia Papadat-Bengescu, *Concert din muzică de Bach*, ESPLA, [1957], M. Sadoveanu, *Opere*, I, Ed. Fundațiilor, 1940, *Dumbrava minunată*, București f.a., Liviu Rebreanu, *Răscoala*, ESPLA, 1957, T. Arghezi, *Cîntare omului*, ESPLA, 1956, *Tablete de cronicar*, ESPLA, 1960, Zaharia Stancu, *Descult*, ESPLA, 1962, Titus Popovici, *Setea*, ESPLA, 1958, Francisc Munteanu, *Statuile nu rîd niciodată*, Ed. tineretului, 1961, Eugen Barbu, *Prînzul de duminică*, Ed. pentru literatură, 1962, I. Vinea, *Lunaticii*, EPL, 1965, Ștefan Bănulescu, *Iarna bărbajilor*, EPL, 1965.

Pentru stilul publicistic am excerptat materiale din articolele aparținînd următoarelor publicații :

Știința, *România liberă*, *Contemporanul*, *Luceafărul*, *Informația Bucureștiului*, *Gazeta învățămîntului*, *Magazinul*, *Flacăra*, *Săptămîna culturală a Capitalei*, *Programul de radio și televiziune*, *Urzica*, *Almanahul turistic ONT 1962*.

În unele cazuri, am apelat la fragmente din următoarele texte aparținînd stilului științific :

I. Simionescu, *Flora României*, Ed. Fundațiilor, 1939, (colectiv), *Fauna Republicii Populare Române*, I, Ed. Academiei RPR, 1951, V. Babeș, *Opere alese*, I, Ed. Academiei RPR, 1954, Raluca Ripan, Ervin Popper, Candin Liteanu, *Chimie analitică cantitativă*, Ed. tehnică, 1957, E. Beral, M. Zapan, *Chimie anorganică*, Ed. tehnică, 1955, Cornel Irimie, *Portul popular din Țara Oltului*, ESPLA, f.a., Al. Cișman, *Fizica generală*, II, București, 1957, G. Ionescu-Șișești, Irimie Stancu, *Agrotehnica*, Ed. agrosilvică de stat, 1958, (colectiv), *Principii de drept*, Ed. științifică, 1959, (colectiv), *Monografia geografică a Republicii Populare Române*, Ed. Academiei RPR, 1960, A. Kreindler, *Nevroza astenică*, București, 1960, S. Schächter, F. Munteanu, *Aplicațiile mașinilor electronice de calcul*, București, 1960,

Almanahul sănătății, Ed. medicală, 1962, V. Pârvan; *Getica*, București, 1962.

În general, exemplele au totalizat cîte 10 000 (zece mii) de foneme. Am considerat eşantioanele suficient de mari, pentru a putea trage concluzii adecvate asupra colectivelor de fapte prezentate. În cazuri speciale, cînd am efectuat sondaje de probă, verificări etc., am menționat numărul de foneme la care ne-am raportat.

0.7.2. Unitățile fonologice cu care am lucrat sînt fonemele din inventarul stabilit de Em. Vasiliu, în *Fonologia limbii române*. Am preferat această descriere, între altele, pentru că este mai operantă din punct de vedere statistic.

Sistemul vocalic al limbii române cuprinde vocalele : /a, e, i, ă, î, o, u/ cu alofonele lor;

sistemul consonantic cuprinde consoanele : /p, b, f, v, t, d, s, z, ț, Ț, 3, ċ, ġ, k, g, h, l, r, m, n/ cu alofonele lor.

1. TIPURI DE STRUCTURI CANTITATIVE

1.1.1. Fonemele românești

Cu toate că ne interesează fonemele românești mai ales în poziții determinate și, legat de aceasta, în succesiunea lor, considerăm că este necesar să știm și care este, în general, situația fonemelor în limbă¹.

Este vorba de caracteristicile cantitative ale fonemelor în indiferent ce poziție.

Vom obține o primă imagine a sistemului fonologic românesc apreciind raporturile cantitative dintre foneme.

Raportul: calitativ/cantitativ în stabilirea inventarului fonematic poate fi interpretat astfel:

a) primul nivel de cercetare se caracterizează *matematic* prin diviziunea elementelor într-un grup cu probabilități pozitive de ocurență și unul cu probabilități zero. Prin această operație se elimină materialul care nu aparține idiomului studiat (limbii) și se iau în evidență toate fonemele proprii lui;

b) Se înregistrează toate fonemele din idiomul respectiv indicându-se probabilitatea fiecărui fonem și se formulează reguli generale care rezultă din distribuția probabilistică obținută².

Investigarea textelor pune în lumină următoarele procente³ (cu erorile standard respective):

¹ Pentru limba română s-a efectuat recent o cercetare pe un număr foarte mare de elemente sintagmatice. Nu ne putem referi însă la rezultatele obținute de autor, dat fiind că e vorba de o statistică pe *litere* — text. Vezi Octavian Tocaci, *Unele date statistice privind frecvența literelor și digramelor în limba (scrisă) contemporană*.

² W.J. Plath, *The Relative Frequency of English Consonantal Patterns*, p. 67—87.

³ Aici și mai departe, atunci când totalul procentelor nu însumează 100 (o sută), înseamnă că am lucrat cu aproximație prin lipsă; am luat în considerație două zecimale (când am notat numai o zecimală, aceasta indică faptul că a doua era zero).

Fonem	%	Eroare-Standard %	Fonem	%	Eroare-Standard %
a	8,8	0,52	l	3,1	0,32
ă	5,7	0,42	m	4,09	0,36
b	0,85	0,14	n	7,6	0,48
k	4,3	0,35	o	5,4	0,41
č	1,2	0,20	p	2,3	0,26
d	3,6	0,35	r	7,5	0,48
e	10,9	0,55	s	4,7	0,39
f	1,1	0,17	ș	1,2	0,20
g	0,9	0,14	t	5,42	0,41
ğ	0,13	0,06	ț	1,02	0,17
h	0,2	0,06	u	7,1	0,47
i	8,1	0,50	v	1,5	0,22
î	3,4	0,32	z	0,5	0,10
3	0,13	0,06			

Clasificînd rezultatele obținute conform cu rangurile fonemelor (deci, în funcție de frecvențe, după locurile în ordine descrescătoare, pe care le ocupă fonemele în tabel) obținem imaginea grafică de la p. 29.

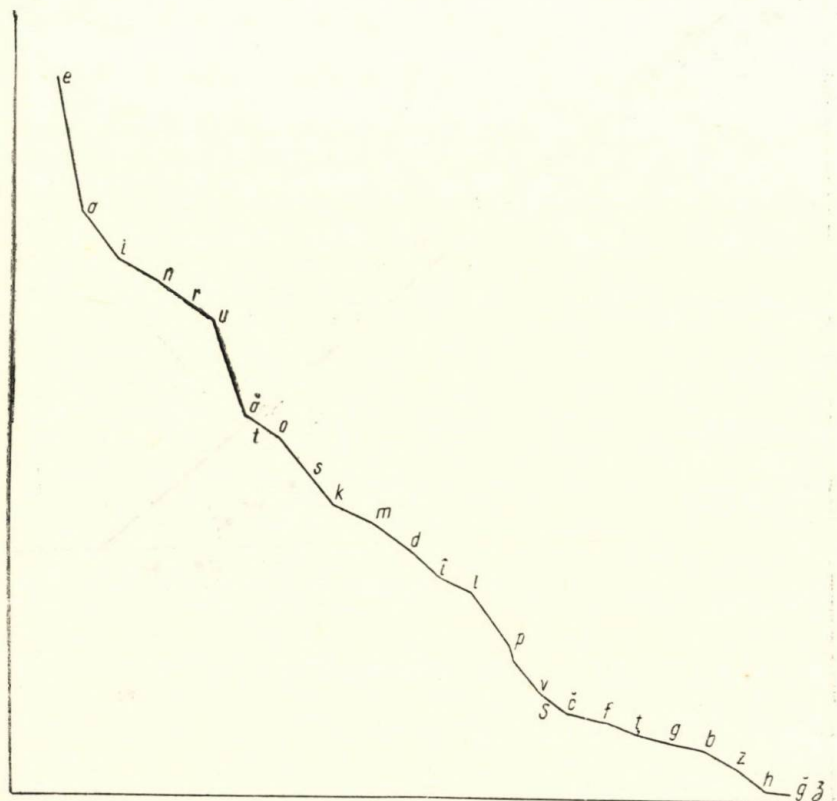
În reprezentare logaritmică, imaginea grafică este cea indicată la p. 30.

Pentru frecvențele de la 10,9 (/e/) pînă la 0,9 (/g/) este verificată relația semilogaritmică dată de Edmond Nicolau (în *Langage et stratégie*, p. 153—177). Pentru valorile sub 1 apar abateri importante care pot fi approximate prin alt segment de dreaptă.

Între fonemul cel mai frecvent (e) și cele mai puțin frecvente (ğ, 3) diferența este de 9,87.

Variația de la un procent de ocurență la altul este însă mică, diferențele nedepășind, în general, cifrele 1—2; ea este ilustrată prin panta lină a curbei pe care se înscriu în ordine fonemele.

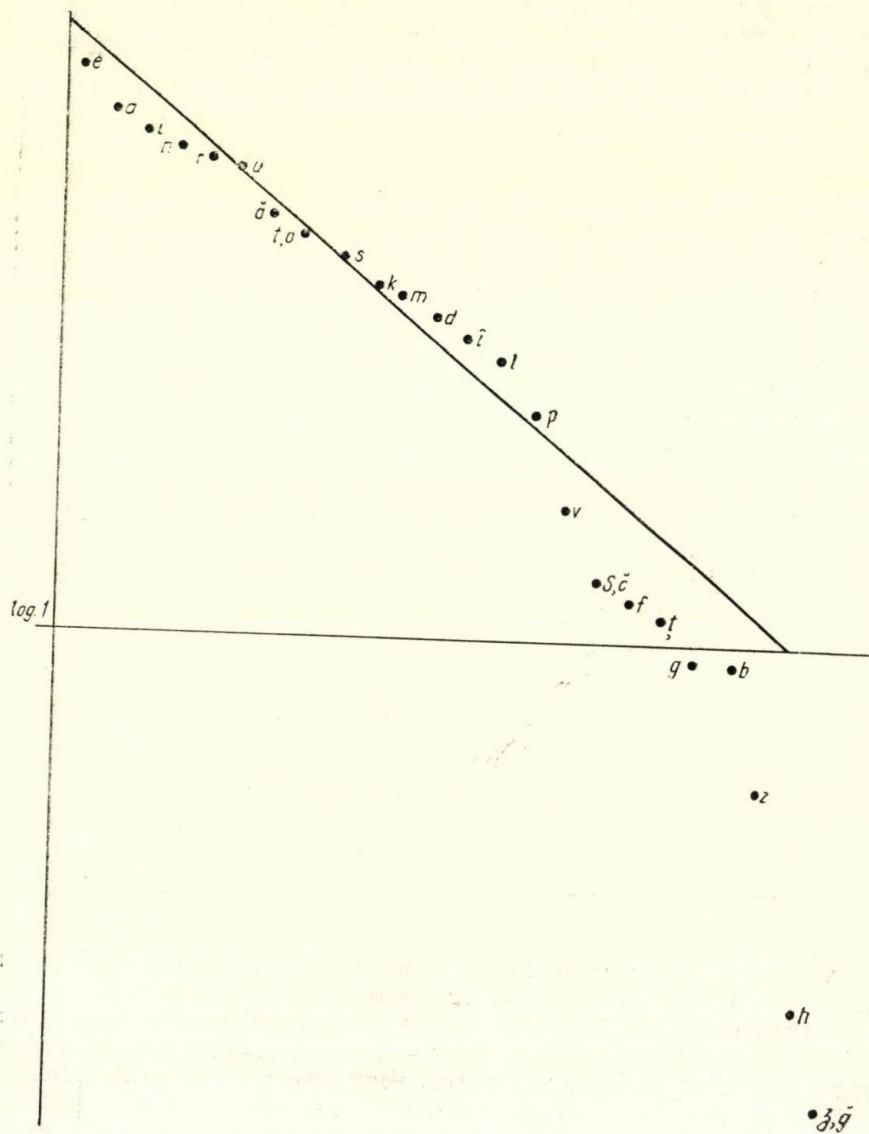
În ansamblul tabelului — la cele 27 de ranguri — vocalele, avînd frecvențe mari, ocupă ranguri mici (primele trei — *e*, *a*, *i*, — apoi rangul șase și șapte *u*, *ă*, rangul nouă — *o* — și, în fine, rangul



paisprezece — *î*), aproape nedepășind (cu *î*, vocala cea mai puțin frecventă) prima jumătate a tabelului.

Pentru consoane se mai pot stabili următoarele raporturi:

Cele mai frecvente consoane — *n*, *r*, *t*, — figurează în prima treime din tabel, la rangurile patru, cinci, șase; *n* este primul fonem care „întrerupe” succesiunea vocalelor; apoi *s*, *k*, *m*, *d* se înscriu la rangurile zece — treisprezece pînă la jumătatea tabelului, unde



apare vocala *i*, care împarte în două şirul descrescător din punct de vedere cantitativ al fonemelor româneşti.

În general, consoanele surde sînt mai frecvente decît perechile lor sonore (pe de o parte, per total, surdele însumînd 20,02 — faţă de sonore care totalizează 7,80 — deci de circa trei ori mai mult; pe de altă parte raportul se menţine şi pentru fiecare pereche surdă/sonoră în parte, cu excepţia perechii consonantice *f/v*, unde diferenţa nu este semnificativă — 0,5 — deci subunitară).

De asemenea, oclusivele dentale (în total 9,2) sînt mult mai frecvente decît cele labiale (în total 3,25) şi chiar decît cele velare (în total, 4,7)⁴.

Raportul dintre vocale şi consoane este următorul :

vocale = 49,4

consoane = 49,74

Studiile fonostatistice efectuate anterior pentru limba română⁵ înregistrează următoarele rezultate :

foneme	I	II	III
vocale	42,7	45,16	47,26
consoane	[57,3]	54,84	52,74

⁴ Această configuraţie corespunde, în linii mari, cu raporturile stabilite de G.K. Zipf — pentru consoanele surde/sonore — şi cu observaţiile ulterioare ale lui W. Mañezak, *Fréquence de l'emploi des occlusives labiales, dentales et vélaires*, p. 208—214. Pentru o comparaţie absolut sumară, cf. şi datele stabilite pentru italiană şi reproduse pentru alte limbi de Jiří Kramský, *A Quantitative Phonemic Analysis of Italian Mono-, Di-, and Trisyllabic Words*, p. 139, privind „ierarhizarea” consoanelor celor mai frecvente :

lb. germană : p, l, s, j, ŋ;
 „ engleză : t, l, s, k, n;
 „ cehă : p, v, n, l, k;
 „ italiană : r, t, n, s, l.

Se observă şi aici că cele mai multe sînt dentale.

⁵ Aceste cercetări aparţin următorilor autori :

I = I. Şiadbei, *Din estetica limbii române : Proporţia fonemelor*, p. 18. Notăm datele lui I. Şiadbei doar pentru comparaţie, avînd în vedere că autorul a efectuat numai sondaje (10 texte × 100 de foneme), în care, de altfel, au fost însumate, ca elemente cu pondere egală, proprietăţi numerice caracterizînd graiuri şi epoci diferite. Între [] completăm procentele consoanelor, deduse de noi în raport cu cele ale vocalelor.

II = D. Macrea, *Despre fizionomia fonetică a limbii române*, p. 33. Studiul este o reeditare (cu modificări) a lucrării *Frecvenţa fonemelor în limba română*, în *Dacoromania*, X, 1941—1943, p. 39—49.

III = M. Grumăzescu, *Studiu asupra foneticii statistice a limbii române*, p. 1432.

Raportul vocale/consoane este considerat un parametru statistic important în caracterizarea limbilor. Consemnăm o serie de date comparative, pentru a vedea care este situația limbii române între alte limbi ⁶.

Limba	Vocale	Consoane
franceză	43,36	56,64
italiană	47,74	52,27
germană	38,86 (40)	61,14
engleză-americană	37,4 (36,11)	62,6 (62,97)
suedeză	41	59
sîrbocroată	46,47	53,53
cehă	40,73	59,13
slovacă	43,82	56,18
neogreacă	46,61	53,59
maghiară	41,91	58,09 ¹⁾

Potrivit acestor rezultate, limba română se apropie cel mai mult de limba italiană; se deosebește cel mai mult de limbile germanice.

⁶ Datele comparative au fost extrase din: pentru limbile: franceză, italiană, germană [prima cifră], sîrbocroată, neogreacă — A. Meillet, A. Vaillant, *Grammaire de la langue serbo-croate*, Paris, 1924, p. 16.

Pentru limbile: germană [a doua cifră] — P. Menzerath, W. Meyer-Eppler, *Sprachtypologische Untersuchungen*, p. 92; engleză-americană [prima cifră] — Rebecca E. Hayden, *The Relative Frequency of Phonemes in General-American English*, p. 217 — 223, [a doua cifră] — A. Hood Roberts, *A Statistical Linguistics Analysis of American English*, p. 39; suedeză — Mary Weiss, *Über die Relative Häufigkeit der Phoneme des Schwedischen*, p. 43; cehă — G. Herdan, *Language as Choice and Chance*, p. 71; maghiară — Edit Vértess, *Statistische Untersuchungen über den Phonetischen Aufbau der Ungarischen Sprache*, p. 136; slovacă — Ján Bosák, *Frequency of Phonemes and Letters in Slovak and Numerical Expression of Some Phonemic Relation*, p. 120 — 136.

Între limbile care figurează în tabelul nostru, nu sînt însă diferențe cantitative *foarte* mari — privind întreaga clasă a vocalelor și respectiv întreaga clasă a consoanelor. Ne putem aștepta, deci, la diferențe mai sensibile între indicii numerici ai distribuțiilor diverselor foneme. Din asemenea deosebiri ar rezulta specificul „eufonic” al fiecărei limbi în parte.

Rezultatele expuse pînă acum sînt valabile atîta timp cît urmarim în texte unități izolate.

De îndată însă ce ne propunem să analizăm un text, avînd în vedere atît segmentele simple (vocală, consoană) cît și segmentele complexe (vocală + vocală; vocală + vocală + vocală; consoană + + consoană etc.), caracteristicile cantitative se modifică.

Un sondaj pe număr limitat de foneme (2000 de foneme) ne oferă următorii indici :

Segmente simple V :

a	6,1
ă	5,6
e	12,8
i	7,5
î	2,3
o	3,8
u	6,6

Segmente complexe VV

ae	0,3	eu	0,15	oa	0,6
ai	0,3			oi	0,05
au	0,2				
ău	0,05	ia	0,5	ua	0,1
		ie	0,2	uă	0,05
		io	0,05	ui	0,4
		iu	0,2	uo	0,05
ea	1,01				
ei	0,1	ii	0,05		

Segmente complexe VVV

ăie	0,05
eau	0,05
eia	0,05
iau	0,1
oie	0,05
ouă	0,05
uie	0,1

Segmente simple C

b	0,95	m	5,3
k	3,8	n	4,6
č	1,9	p	1,6
d	3,8	r	3,7
f	1,01	s	2,8
g	0,4	ș	1,2
ğ	0,2	t	3,9
h	0,05	ț	1,01
3	0,15	v	1,6
l	3,6	z	1,2

Segmente complexe CC

br	0,05	mb	0,05	pr	0,5		
bs	0,05	mn	0,2	pt	0,4	șk	0,05
kl	0,1	mp	0,05			șn	0,05
kr	0,1	mt	0,05	rb	0,1	șt	0,30
kt	0,05	mț	0,05	rk	0,15		
		mv	0,05	rč	0,05	tr	0,2
dr	0,3	nc	0,2	rd	0,05	vr	0,1
dv	0,15	nd	0,2	rm	0,15	zb	0,05

Segmente complexe CCC

		nf	0,2	rn	0,05		
fl	0,05	ng	0,25	rs	0,05	ltë	0,05
fr	0,1			rj	0,05	mpl	0,05
gr	0,1	nğ	0,05	rt	0,05	ndr	0,05
		ns	0,05	rv	0,05	ngr	0,1
						nst	0,05
lë	0,05	nj	0,05	sk	0,15	ntr	0,25
ln	0,05	nt	0,6	sf	0,05	skr	0,05
ls	0,05	nț	0,5	sm	0,05	spr	0,1
tl	0,05	nv	0,05	sp	0,15	str	0,4
lv	0,05	nz	0,05	st	0,8		

Felul cum se „ierarhizează” de astă dată diversele elemente fonice este indicat la p. 36.

Se observă că fonemele sînt „introduse” pe rînd în text.

Primele 5 ranguri aparțin numai vocalelor; rangurile 6—22 sînt ocupate și de vocale dar și de consoane (după rangul 15, de fapt, nu mai apar vocale); la rangul 23 intervine primul grup vocalic cu două vocale (*ea*); iar la rangul 25 primul grup consonantic cu două consoane (*st*); dintre grupurile cu trei elemente omogene, apar mai întîi consoanele, la rangul 34 (*str*), și apoi vocalele, la rangul 55 (*iau, uie*). După acestea, la rangul 56, nu ar mai figura decît o consoană (*h*) și o serie de grupuri vocalice și consonantice.

Remarcăm că, într-o investigație pe material mai restrîns, nu apar de loc grupurile fie vocalice, fie consonantice, cu peste trei elemente, altfel posibile în limba română.

Considerarea tuturor elementelor constitutive — simple sau complexe — „modifică” frecvențele înregistrate numai pentru segmentele vocalice și consonantice simple și, implicit, aranjarea respectivă după rang (vezi p. 37).

Intervine de astă dată gradul de combinabilitate a elementelor, aderența sau inaderența lor.

O asemenea descriere a limbii este necesară, tocmai pentru înțelegerea completă a structurii unui mesaj și pentru a putea stabili condițiile optime de transmitere a mesajului cu un minim de material.

Rang	Segment					
1	e-12,8 i u a ă-5,6					
6		m-5,3 n t d, o r k l s i c p, v ʃ, z-1,2 t, f-1,01 b-0,95				
21-23			ea-1,01			
25				st-0,8		
31-34	g		oa; nt ia ui-0,4 ai, ae au, ie	nț, pr pt-0,4, șt, dr mn, nč, nd, nf, ng, tr	str-0,4 ntr	
50-54	z 3-0,1		eu ei, ua-0,1	ngr, spr-0,1		iau, uie-0,1
.	.	...	...	...	...	...

Pentru caracterizarea informației pe care ne-o dă o secvență de foneme, este important să știm, de exemplu, pe de o parte, dacă e mai probabil ca după o vocală să urmeze tot o vocală etc. Pe de altă parte, însă, aproximarea devine mai fină, în momentul cînd știm precis că unei vocale *e* îi urmează mai probabil *a* (*ea*) decît *i* (*ei*). De asemenea, este util să știm că *e* apare în mai mare măsură necombinat (12,8) decît în toate combinațiile sale (*ea* + *ei* + *eu*)

laolaltă (1,26) — ceea ce rezultă dintr-un calcul comparativ. (Facem această precizare, datorită faptului că pentru *e*, înregistrat numai izolat în raportul vocale-consoane, nu și în grupuri vocalice, caracteristica numerică este 10,9.)

1.1.2. În cercetarea noastră am urmărit, în general, grupurile fonematice omogene, în cadrul cuvîntului (secvențe cuprinzînd numai vocale, sau numai consoane).

Există însă și alte modalități de investigare: de exemplu, fiind propusă dimensiunea unei secvențe (de 2, 3, 4, ... foneme) indiferent de calitatea componentelor ei, să se calculeze cantitatea de informație pe care o oferă ea, fiind date probabilitățile alăturării simbolurilor respective ⁷.

Ne propunem să stabilim gradul de aderență a elementelor care compun segmentele omogene sau heterogene.

Astfel, vom compara frecvența grupurilor omogene de tipul VV, sau CC, cu a grupurilor de tip VC, sau CV.

O investigare a lor pune în lumină următoarele proporții:

$$\begin{aligned} V + C &= 25,6 \\ C + V &= 60,2 \\ V + V &= 4,6 \\ C + C &= 8,9 \\ V + V + V &= 0,06 \\ C + C + C &= 0,5 \end{aligned}$$

Aceste structuri denotă preferința limbii române pentru grupări de elemente de tip diferit ($[V + C] + [C + V] = 85,8$) față de grupările cu elemente omogene ($[V + V] + [C + C] + [V + V + V] + [C + C + C] = 14,06$). Sudura cea mai puternică este cea între grupurile $C + V$ și cea mai slabă între grupurile $V + V$.

Dintre limbile pentru care dispunem de date comparative, situația cea mai asemănătoare o oferă limba rusă ⁸.

⁷ Vezi, de exemplu, pentru un astfel de tip de cercetare, D.S. Lebedev și V.A. Garmaș. *Statističeskij analiz trebkuvennih sočetanij russkogo teksta*, p. 78—80.

⁸ Pentru limba rusă, combinațiile dintre foneme au următoarele probabilități:

$$\begin{aligned} C + V &= 0,7449 \\ C + C &= 0,2551 \\ V + V &= 0,0017 \end{aligned}$$

$V + C = 0,9983$ L.R. Zinder, *Despre probabilitatea lingvistică*, p. 194. Cf. și expunerea lui Gustav Herdan, *The Numerical Expression of Selective Variation in the Vowel-Consonant Sequence in English and Russian*, p. 102—103, unde se demonstrează și că în cazul opus limbii ruse, limba engleză, preferința este pentru grupuri fonematice omogene. În limba maghiară, grupările $C + C$ au frecvența 9,25%, secvențele VC — reprezintă 35,75, iar cele $C + V$, 34,75. Cf. Edit Vertés, *lucr. cit.*, p. 148.

Urmează să caracterizăm cantitativ fonemele românești în diverse poziții, respectiv contexte fonologice.

1.2. Inițiala cuvintelor

Cercetarea *inițialei* cuvîntului oferă o serie de rezultate specifice, dat fiind că în ierarhizarea elementelor componente ale cuvintelor, partea inițială deține un loc deosebit de interesant în ce privește cantitatea de informație conținută în cuvîntul respectiv⁹.

În cele ce urmează, ne referim la segmentele inițiale simple, adică la fonemele care urmează imediat după pauză. Avem în vedere următoarele cazuri: [A] texte literare; [B] texte de limbă vorbită; [C] dicționar.

Pentru a stabili frecvența inițialei în textele de limbă literară am utilizat rezultatele statistice obținute de V. Șuteu¹⁰ pe baza unui material extras din operele a 10 scriitori români. Aceste rezultate referindu-se însă la *litera* inițială, am întreprins un sondaj privitor la fonemele inițiale, acolo unde o literă desemna mai multe foneme ($c = /k/, /č/, g = /g/, /ğ/$), sau invers, mai multe litere reprezentau un singur fonem ($che, chi = /k'/, ghe, ghi = /g'/$, alofonele lui $/k/$ și $/g/$).

Pentru limba vorbită, am stabilit frecvența fonemelor inițiale pe baza unor texte populare¹¹.

Al treilea caz pe care l-am examinat este frecvența fonemelor inițiale într-un dicționar¹².

Trebuie să observăm că, în general, toate fonemele românești pot apărea în poziție inițială:

⁹ R. G. Piotrovski, *Problema locului în cuvînt a elementelor purtătoare de informație*, p. 32—33.

¹⁰ Materialul excerptat de V. Șuteu cuprinde fragmente din operele lui: M. Sadoveanu (*Nicoară Potcoavă*), Ion Agirbiceanu (*Arhanghelii*), I. Al. Brătescu-Voinești (*Două lumi*), Tudor Arghezi (*Versuri*), Mihai Codreanu (*Sonete*), Octavian Goga (*Poezii*), Eusebiu Camilar (*Turmele*), Liviu Rebreanu (*Ion*), Marin Preda (*Moromeții*), Camil Petrescu (*Teatru*); cf. V. Șuteu, *Observații asupra frecvenței cuvintelor în opera unor scriitori români*, p. 419—443.

¹¹ Textele populare au fost culese de la mai multe subiecte și publicate de: Alexandrina Istrătescu, *Texte populare din județul Prahova*, în *Grai și suflet*, III, București, 1927—1928, p. 154 și urm.

¹² *Dicționarul limbii române literare contemporane*, Ed. Acad. R.P.R., vol. I, 1955; vol. II, 1956; vol. III și IV, 1957. N-am luat în considerație literele *k, w, x, y*, pe care le-am considerat simple grafii.

/ak, ăla, basma, kopil, čară, dar, economie, fir, gînd, ănană, haină, imn, în, 3ar, lemn, măr, nor, om, pom, rînd, sare, și, tare, 3ap, unire, vin, zînă/ (*ac, ăla, basma, copil, ceară, dar, economie, fir, gînd, geană, haină, imn, în, jar, lemn, măr, nor, om, pom, rînd, sare, și, tare, 3ap, unire, vin, zînă*).

Fonemele se ierarhizează, în cazurile studiate de noi, în modul următor :

[A] — în textele literare, de la *s* la *ă* :

s(10,1), *k*(8,93), *p*(9,8), *a*(6, 7), *î*(6,5), *d*(5,6), *m*(5,6), *b*(4,5), *n*(4), *t*(3,9), *u*(3,5), *v*(3,2), *l*(3), *g*(2,3), *r*(2,2), *o*(2,5), *č*(2,26), *i*(2,1), *z*(1,8), *h*(1,3), *ș*(1,1), *e*(0,8), *3*(0,7), *ț*(0,7), *ğ*(0,6), *ă*(0,1).

[B] — în textele populare, de la *a* la *ğ* :

a(10,7), *k*(9,1), *p*(7,8), *s*(7,25), *m*(6,2), *e*(5,9), *l*(5), *ș*(4,85), *n*(4,85), *o*(4,2), *f*(4), *b*(3,8), *č*(3,7), *t*(2,4), *u*(2,3), *v*(2,15), *i*(1,8), *g*(1,61), *î*(1,51), *r*(1,04), *ă*(0,47), *h*(0,28), *z*(1), *3*(0,19), *ğ*(0).

[C] — în dicționar, de la *p* la *ă* :

p(11,5), *s*(9,84), *k*(9,80), *a*(6,44), *d*(5,91), *m*(5,84), *t*(5), *r*(4,56), *î*(3,98), *f*(3,52), *i*(3,39), *b*(3,3), *v*(3,14), *n*(2,96), *l*(2,6), *o*(2,57), *e*(2,19), *h*(2,08), *g*(2,027), *z*(1,67), *ș*(1,61), *č*(1,5), *u*(0,35), *ț*(0,98), *3*(0,79), *ğ*(0,6), *ă*(0,02).

Pornind de la aceste date, putem înscrie într-un tabel rezumativ (p. 40) frecvența tuturor fonemelor *inițiale*, în aceeași ordine, pentru a avea o imagine integrală asupra fiecărui fonem, care să ne permită în același timp compararea diverselor sale procente de apariție în texte diferite.

Considerînd situațiile pe care le ilustrează textele de mai sus [A], [B], [C], am înregistrat cele mai mari variații¹³ de frecvență pentru fonemele /*e*/ și /*î*/ (5%) și cele mai mici variații de frecvență pentru fonemele /*ă*, *k*, *č*, *g*, *ğ*, *f*, *3*, *m*, *ț*/ (1% și sub 1%); frecvența celorlalte foneme, la inițială, variază în proporție de 2—4%. În raport cu valoarea absolută a procentelor, variațiile înregistrate pentru aceste cazuri sînt comparabile ca ordin de mărime (astfel, pentru *e* variația maximă este 5%, probabilitatea lui *e* la inițială fiind 5,9% în cazul [B], 0,8% în cazul [A] și 2,19% în cazul [C].

O situație asemănătoare se constată dacă avem în vedere frecvența diferitelor foneme pentru fiecare din cazurile [A], [B],

¹³ Avem în vedere, în aceste cazuri, frecvențele *maxime* ale aceluiași fonem.

[C], unde diferențele maxime sînt 0,1—10,4% ($\check{a}-p$) pentru [C]. De aceea, am completat, prin indicații de rang, comparația dintre probabilitățile pozițiilor inițiale.

Fonem	/A/ (Text literar)	/B/ (Text popular)	/C/ (Dicționar)
a	6,7	10,7	6,44
ă	<i>0,1</i>	0,47	<i>0,02</i>
b	4,5	3,8	4,85
k	8,93	9,1	9,8
č	2,26	3,7	0,5
d	5,7	9,5	5,91
e	0,8	5,9	2,19
f	4	4	3,52
g	2,3	1 61	2,02
ğ	0,6	<i>0</i>	0,5
h	1,3	0,28	2,8
i	2,1	1,80	3,39
î	6,5	1,51	3,98
Ț	0,7	0,19	0,79
l	3	5	2,6
m	5,6	6,2	5,84
n	4	4,85	2,96
o	2,5	4,2	2,57
p	9,8	7,8	11,5
r	2,2	1,04	4,57
s	10,1	7,25	9,84
ș	1,1	4,85	1,61
t	3,9	2,4	5
ț	0,7	0,19	0,89
u	3,5	2,3	0,35
v	3,2	2,15	3,14
z	1,8	1	1,67

cifrele aldine = frecvență maximă
cifrele cursive = „ minimă

Pentru a ilustra mai bine locul specific al inițialei în organizarea cuvîntului, am adăugat o serie de informații cu privire la frecvența tuturor fonemelor românești în *toate* pozițiile, în textele literare și în textele populare.

Fonemele se organizează, în aceste cazuri, [D] în textele literare¹⁴, de la *e* la *z* (vezi cap. 1, p. 28), [E] în textele populare¹⁵, de la *a* la *h* :

a(13,07), *e*(10,28), *i*(7,59), *u*(7,59), *n*(6,42), *k*(5,79), *t*(5,76), *ă*(4,7), *r*(4,49), *m*(4,32), *s*(4,04), *l*(4,04), *o*(3,94), *p*(2,5), *î*(2,4), *ș*(1,95), *f*(1,87), *b*(1,72), *v*(1,86), *ț*(0,77), *z*(0,57), *g* (0,55), *ć* (0,2), *ž*(0,19), *ǵ* (0,08), *h*(0,02).

Utilizînd toate datele de mai sus, am alcătuit un tabel recapitulativ (p. 42) al situațiilor [A], [B], [C], [D] și [E] pentru cele 25 de foneme considerate, clasificate după *rang*. Am divizat (arbitrar) acest tabel în 3 secțiuni, fiecare cuprinzînd cîte 9 foneme și am comparat : *a*) probabilitatea de apariție a fonemelor ca inițiale în textele [A], [B], [C] și *b*) abaterea (*écart*) pe care o indică poziția inițială în raport cu probabilitatea de apariție în toate pozițiile pe care le pot avea fonemele respective în limbă.

Secțiunea I

a figurează în această secțiune pentru toate situațiile (atît ca inițială, cît și în toate celelalte poziții) ; *m*, *s*, *k*, *p*, *d*, figurează, de asemenea, în această secțiune, ca inițiale, dar probabilitatea lor în poziție inițială nu concordă cu probabilitatea de apariție în celelalte poziții. Astfel, pentru *k*, *p* cazurile [D] [E] apar în secțiunea a doua a tabelului ; pentru *m* și *s*, cazurile [D] și [E] apar tot în secțiunea II.

Secțiunea II

Fonemul *f* (inclus de noi printr-un artificiu grafic în secțiunea II, pentru cazul [A], cînd se afla la limita dintre secțiunile I și II) figurează la inițială, la această secțiune : *v*, *o*, *i* și *g* și *n* (incluse prin același artificiu grafic), de asemenea. În celelalte situații, pentru *v*, cazul [D] se află tot în secțiunea II, iar [E] se află în secțiunea III ;

¹⁴ Pentru comparații, vezi și Edmond Nicolau, C. Sala, Al. Roceric, *Observații asupra entropiei limbii române*, p. 41.

¹⁵ Cf. Alexandrina Istrățescu, *op. cit.*

[A]	[B]	[C]		[D]	[E]
	(inițială)			(Toate pozițiile)	
1	s	a	o	e	ă
2	k	k	s	ă	e
3	p	a	k	i	i, u
4	a	p	a	n	
[I] 5	i	s	o	r	n
6	o	m	m	u	t
7	m	e	t	ă	ă, d
8	b	l	b	t	
9	f, n	f, n	r	o	r
10			i	s	m
11	t	o	f	m	s, l
12	u	f	i	k	
13	v	b	v	d	o
[II] 14	l	č	n	i	k
15	q	t	l	l	p
16	r	u	o	p	i
17	o	v	e	v	f
18	č	i	h	f, č	
19	i	g	g		b
20	z	i	z	f	v
21	h	r	f	t	t
22	f	z	č	b	z
[III] 23	e	ă	u	g	g
24	ă, t	h	t	z	č
25		ă, t	ă	h	ă
26	q	q	q	q, ă	q
27	ă	q	ă		h
————— Concordanță de rang în toate cazurile ----- Concordanță de rang, numai pentru poziția inițială					

pentru *g*, și [D] și [E] în secțiunea III; pentru *i* și *n* în secțiunea I, iar pentru *o*, cazul [D] se află în secțiunea I și cazul [E] în secțiunea II.

Secțiunea III

Fonemele *z*, *3*, *t*, *g*, apar în această secțiune în toate cazurile; fonemul *ă* în poziție inițială; cazurile [D] și [E] pentru *ă* apar în secțiunea I.

Se observă deci că 17 foneme din 27 (*s*, *k*, *p*, *a*, *d*, *m*, *f*, *n*, *v*, *g*, *o*, *i*, *z*, *3*, *t*, *g*, *ă*) au o apariție regulată în poziție inițială (chiar atunci când intervin *diferențe de frecvență*, fonemele respective ocupă *ranguri apropiate* în toate cazurile examinate de noi); celelalte 11 foneme (*î*, *b*, *t*, *u*, *r*, *č*, *h*, *ș*, *e*) se repartizează în mod neregulat.

Datele privitoare la *frecvență* și *rang*, conduc la clasificarea fonemelor în trei categorii:

a) cu frecvență mare și cu rang inferior în toate cazurile ([A], [B], [C], [D], [E]);

b) cu frecvență mare și cu rang inferior în poziție inițială (cazurile [A], [B], [C], constituind o abatere de la situația din limbă în general ([D], [E]);

c) cu frecvență și rang accidentale și neconcordând cu nici o lege *aparentă* de repartiție.

Variația frecvențelor de la un caz la altul, în poziție inițială, permite câteva observații:

În general, gruparea unui mare număr de cuvinte la aceeași inițială se explică prin elemente din domeniul foneticii și al formării cuvintelor¹⁶. Trebuie să se țină seama, de altfel, și de corelația dintre frecvența de apariție a unui fonem și frecvența cuvintelor care conțin acest fonem. Este firesc ca atunci când un fonem constituie inițiala mai multor cuvinte foarte frecvente în limbă, procentele care redau apariția sa într-un text oarecare să fie mult mai mari decât în cazul unui fonem care ocupă locul inițial în cuvintele puțin frecvente ale limbii respective¹⁷.

Pentru limba română se pot stabili, în acest sens, concordante, prin compararea inițialelor din fondul principal de cuvinte al limbii

¹⁶ Al. Graur, *Încercare asupra fondului principal lexical al limbii române*, p. 55–57.

¹⁷ Pentru raportul cuvint-literă, vezi și I.M. Iaglom, R.L. Dobrușin, A.M. Iaglom, *Teoria informației și lingvistica*, p. 28.

române¹⁸ cu probabilitatea inițialelor din tabelul alcătuit de noi pentru pozițiile inițiale. Se observă că fonemele cele mai frecvente — ca inițiale — în fondul principal (*k, s, p, m, a*) apar totdeauna în prima secțiune a tabelului (care indică *rangul* fonemelor inițiale), în timp ce fonemele mai puțin frecvente în fondul principal apar în secțiunea a doua a tabelului respectiv și ocupînd un rang mijlociu în tabloul de frecvențe stabilit în funcție de fondul principal (*n, l, g, o*). N-am fost în măsură să stabilim și alte paralele, printre altele pentru că stabilirea fondului principal nu se orientează numai după criteriul frecvenței¹⁹.

Studiul statistic al inițialei pune în lumină frecvențe diferite ale fonemelor în această poziție, față de frecvențele fonemelor în toate pozițiile (inițială, medială, finală). În acest fel, se ajunge la o cantitate de informație specifică poziției inițiale a fonemelor²⁰.

1.3.1 Grupuri vocalice

Pentru această descriere am pornit de la mulțimea de vocale $V = \{a_1, a_2, \dots, a_n\}$ unde a_1, a_2 etc. simbolizează toate vocalele limbii române și de la mulțimea de consoane $C = \{b_1, b_2, \dots, b_n\}$ unde b_1, b_2 etc. simbolizează toate consoanele limbii române. Am urmărit combinațiile posibile între vocalele $a_1, a_2, \dots, a_n$, în segmente complexe. Începem prin a descrie vocalele grupate cîte două și distribuția lor în contextele consonantice $b_1, b_2, \dots, b_n$, la care am adăugat posibilitatea: pauză (#).

Prin *poziție distribuțională* vom înțelege locul ocupat într-un șir de foneme de un grup de foneme (în cazul nostru, vocale) definit în relație cu un context specificat; *contextul* va apărea sub forma altor foneme (în cazul nostru, consoane) care precedă sau urmează grupul în șirul respectiv. Nu ne interesează norma homo-, respectiv hetero-silabică a segmentelor, ci numai alăturarea fonemelor. (Așadar, noi luăm în considerație secvența *ea*, de exemplu, fie că ea se realizează ca [ea] în [bea], fie ca suma fonemelor [e + a] în

¹⁸ Al. Graur, *op. cit.*, p. 56.

¹⁹ Al. Graur, *op. cit.*, p. 56.

²⁰ O descriere *pozițională*, (inițială, medială, finală) a fonemelor românești a întreprins și N. Saramandu, *Sur le rendement fonctionnel des types de structures phonématiques en roumain*. Materialul pe care se bazează lucrarea sa cuprinde cuvintele-titlu din dicționar (DLRM).

[kreare]. Același criteriu va funcționa și în descrierea grupurilor consonantice. *Vezi mai departe* ²¹.)

1.3.2. Cercetarea secvențelor formate din vocală + vocală pune în evidență mai multe tipuri de grupuri vocalice. Definim ca *grupuri vocalice simetrice* grupurile în care, permutînd cele două elemente între ele, obținem grupuri proprii limbii române.

Dintre cele 10000 de grupuri vocalice cu două elemente, pe care le-am inventariat, circa 94,5% sînt grupuri simetrice (considerînd toate pozițiile: inițială, medială, finală).

Au rezultat următoarele grupuri vocalice simetrice, distincte: *ae/ea, ai/ia, ao/oa, au/ua, ău/uă, ei/ie, eo/oe, eu/ue, io/oi, iu/ui, îu/uî, ou/uo*.

Grupurile de mai sus diferă între ele, atît prin numărul total de apariții, deci prin frecvență, cît și prin posibilitatea de distribuire într-un număr mai mare sau mai mic de contexte diferite. Stabilind aceste contexte, trecem de fapt la descrierea structurii frecvențelor fonematice, ceea ce înseamnă inventarierea seriilor care precedă și a celor care urmează un fonem, precum și studiul lor cantitativ ²².

Această situație se reflectă în tabelul următor:

Grup vocalic	%	Grup vocalic	%
ae	0,7	ea	25,8
ai	8,7	ia	7,1
ao	0,01	oa	8,8
au	5,1	ua	1,2
ău	1,3	uă	0,3
ei	6,6	ie	9,1
eo	1,1	oe	0,07
eu	2,1	ue	0,2
io	2,9	oi	2,10
iu	3,5	ui	10,1
îu	0,05	uî	0,05
ou	0,4	uo	0,05

²¹ De aceea, față de sistemul propus de Em. Vasiliu, *lucr. cit.*, am introdus și o simplificare în transcrierea fonologică, și anume, am renunțat la marcarea joncturii silabice.

²² Cf. Frank Harary and Herbert H. Paper, *Toward a General Calculus of Phonemic Distribution*, p. 147—148.

Din acest tabel rezultă că grupurile cele mai frecvente sînt : *ea* (25,8), *ui* (10,1) ; grupurile cel mai puțin frecvente sînt : *ao* (0,01), *iū*, *uî* (0,05) :

Lista completă²³ a contextelor consonantice în care se distribuie grupurile vocalice (cu două elemente) simetrice este următoarea :
ae : # -(r, v), m-str, v-r

ea : #-#, b-(#, g, m, t), br-z, kl-#, km-l, kr-(#, d, r, st, t, ț, z), ks-z, kt-z, d-(#, l, m, s, sk), dr-pt, f-#, fl-k, ft-#, gl-z, gn-sk, gr-(b, ț), hm-, 3-(#, sk, z), l-(#, f, g, s, sk, ț, z), lb-sk, lm-#, ln-(m, sk), m-(#, rg, sk), mn-(#, v, z), mpl-#, n-(#, gr, l, m, sk, țt, t, ț, z), nd-(#, sk), ndr-pt, ngl-z, n3-z, ns-mn, nt-(#, z), nț-(#, z), ntr-g, p-(#, sk, z), pl-k, pr-(#, 3m, l), ps-#, pt-#, r-(#, l, kț, m, n, s, sk, str, t, z), rb-(#, sk), rm-(#, m), rn-#, rp-#, rs-z, rt-#, rț-sk, rv-#, s-(#, l, m, r, rb, sk, z), st-(#, g), str-z, ț-(#, l), țt-(#, pt), t-(#, l, m, sk, tr, z), tr-(b, kz), ț-(#, l, m, n, p, sk), v-(#, k, m, sk), vr-(#, m, z), z-(#, l, m, z)

ai : #-(#, b, č, d), k-#, km-#, d-#, fl-#, gr-#, h-d, m-(#, k, d), n-(b, l, nt, v), rh-z, sk-v, sp-m, st-#, t-n, tr-#

ia : #-(#, r, rb, rn, s, t), b-(#, t), ks-#, kț-#, č-#, d-(#, g, gr, l, m, t, ț), f-(nt, r, t), l-(#, nt, r, rd), m-(t, z), mb-(nt, nț), n-(#, k, t), nkț-#, nd-l, ns-#, nț-(#, l, t, z), p-(#, n, rd, t, tr, ț), r-(#, l, n, nt, s, ț, t, ț, z), rğ-#, rt-#, rz-(s, t), s-#, skr-#, st-(#, l), ț-(#, t), țt-(#, m), tr-rh, ț-(#, r), v-(l, t, ț), z-(#, r), zm-rd.

ao : d-s

oa : #-(l, m, r, rb, s, sp), b-(l, r), bl-n, k-(č, d, 3, m, n, ps, s, st, t), kt-r, č-r, d-(b, g, mn, r, rm), f-(m, n, rt), fl-r, g-l, gl-t, gr-(p, s, z), 3-(k, č, s), 3l-č, l-(n, r, s), m-(l, r, s, rt), n-(pt, r, s, sk, t, str, țt), nd-k, nt-(rk, rč, rs), p-(r, rt, t), pr-(p, sp, st), pt-s, r-(d, n, s, t), rn-s, rs-n, rv-s, s-(kr, r), sk-t, st-(l, r, s), str-n, ț-(pt, r), țk-l, t-(mn, n, r, rn, t), tr-n, v-(k, č, r, str), zb-r, zv-r

au : #(#, ks, d, g, r, t, z), k-(č, t, ț, z), č-#, d-(#, g), fl-#, g-#, l-#, n-(#, fr, t), nk-#, nğ-#, n3-#, nt-#, p-z, pl-d, r-#, rl-#, rt-#, s-#, sk-n, st-(#r), t-#

ua : k-ț, kt-(#, l, t), d-l, l-(#, s, t, ț), n-(#, l, nț, r, t, ț), nt-l, s-v, t-(r, ț), z-l

²³ În această listă am introdus numai exemple înregistrate de noi în textele pe care le-am parcurs, făcînd deci abstracție de alte contexte cunoscute în limbă. De exemplu : *ae* : k-r (*caer*), *ai* : k-l (*cailor*), kr-# (*crai*), p-# (*pai*), vr-st (*vraște*), o : h-s (*haos*), p-s (*repaos*), *ea* : d-z (*obsedează*), ou : b-# (*bou*), ui : k-# (*cui*) etc.

ău : b-(#, t), k-t, d-g, f-r, l-d, n-ntr, r-(#, l), rn-ț, s-#, t-#, z-#
uă : l-(#, m), n-(#, m)
ei : #-#, kl-(#, k, l), č-(#, l), d-#, gr-#, l-#, lb-#, lt-#, m-(#, l), mn-#, n-#, nt-#, nț-#, p-#, r-#, rm-#, rp-#, rs-#, rt-#, s-#, sl-ț, st-#, t-(#, n), tr-#, ț-#, v-#, vr-#, z-#
ie : #-(d, ft, p, rb, s, ș), b-(#, kt, t), kl-nt, ks-#, kț-#, č-#, d-(#, nț, r, z), dr-r, f-(#, k, r, rb, rt), ģ-#, l-(#, f, r), lm-#, m(#, l, r, z), mbr-#, mn-#, nkț-#, nț-(#, r), nșt-nt, p-(#, d, l, nt, pt, r, rd, s, tr), pr-(#, t), pț-#, r-(#, nt, nț, r, t, z), rğ-#, rm-r, rs-#, rt-#, rz-#, s-#, skr-#, str-#, ș-#, st-#, t-(#, r), ț-(#, t), v-(#, rm, ț), z-#, zm-rd
eo : d-(k, d, s), l-p, n-r, pr-(k, t), t-r, vr-#
oe : k-(f, r), p-t
eu : #-#, kl-(#, l), č-#, gr-(#, t), l-#, m-#, mpr-n, n-r, p-#, pn-m, r-(#, m, ș), t-r, vr-n, z-(#, l)
ue : d-l, l-t, n-#, nfl-nț, ș-t
io : #-n, b-l, kț-n, d-(#, kr, d), f-r, 3d-s, l-n, n-s, nkț-n, nf-r, nț-(n, ș), p-s, pț-n, r-(d, r, s, z), rd-l, rl-nț, rn-n, s-n, sf-s, tr-d, ț-(n, s), v-(n, r), z-n
oi : #-#, k-nč, kr-t, d-(#, l, n), f-#, g-sm, 3-#, m-#, n-(#, l), nd-(#, t), p-#, pl-#, r-(#, k, č, sm), t-#, v-(#, nț, s, t), zb-#
iu : #-(b, t, ț), kț-n, k-#, d-(#, l, m), f-(#, l), l-(#, l), n-(#, l, r), nkț-n, ns-n, p-(#, l), pr-#, pț-n, r-(#, l), rs-n, rz-(#, n), s-n, skr-#, st-(#, n), șt-(#, t), ț-(#, l, n), v-(#, n), z-(#, n)
ui : #-(m, t, ț), b-(#, nț, t), k-(b, nd, nț, s, t, v) d-(#, nț, t), f-s, fl-(d, z), z-(#, nd, t), 3d-(#, t), l-(#, s), lt-#, n-(#, s, t), nstr-(r, t), nț-(#, r, t), p-(#, m, r, t, z), pt-t, r-(#, n, nț, s, t), rk-t, rz-#, sk-(#, p), sp-#, st-#, str-#, șn-(#, t, ț), t-(#, r, t, ț), ț-(#, k, nd, t), v-#, z-#
îu : br-l, r-(l, r)
uî : kt-nd, l-nd, t-nd
ou : bl-(#, r), k-l, d-#, n-(#, l), r-(#, l)
uo : kt-s, d-#, st-s.

Pentru a putea aprecia care grupuri vocalice au distribuția cea mai largă și care se distribuie în mai puține contexte, am stabilit, în tabelul de la p. 48, pentru fiecare grup vocalic, numărul de contexte în care a fost atestat (în acest caz lucrăm deci cu cifre absolute și nu cu procente).

Exemple : /aer, aevea, maestru, vaer, ea, bea, pribeag beam, leat, breaz, stikleă, tokmeală, akrea, kreadă, kreare, kreată, kreastă, kreață, lukrează, fikșează, piktează, vedea, deal, vedeam, deasă, pîndească,

dreaptă, kafea, fleak, pofta, reglează, răstignească, agreabil, greață, duhnea, prășea, slușească, rușează, kalea, leafa, leagă, aleasă, pilească, ațeleași, mulează, albească, kulmea, întâlneam, întâlnească, mea, near-

Grup vocalic	Repartiție la nr. de contexte	Grup vocalic	Repartiție la nr. de contexte
ae	4	ea	141
ai	23	ia	79
ao	1	oa	92
au	34	ua	20
ău	13	uă	4
ei	33	ie	75
eo	9	oe	3
eu	19	ue	5
io	31	oi	26
iu	39	ui	64
iu	3	iî	3
ou	8	uo	3

gă, primească, domnea, dumneavoastră, semnează, umplea, puneă, neagră, černeală, spuneam, kimpenească, čineaști, alineat, dimineața, planează, pîndea, gîndească, îndreaptă, 3onglează, aranșează, înseamnă, neant, kontează, mărunțea, desființează, întreagă, rupea, lipsească, pompează, pleacă, prea, preașmă, popreală, lipsea, proptea, rea, real, reakție, akream, podgorean, mireasă, kučerească, fereastră, șireată, virează, sorbea, vorbească, kîrmea, dormeam, urnea, kîrpea, eksersează, kîrtea, dezmořtească, servea, kosea, oboseală, țesem, seară, searbăd, sosească, visează, stea, steag, păstrează, tușea, nădușeală, ku-noștea, așteaptă, putea, beteală, teamă, rotească, teatru, viteaz, treabă, treacă, treaz, cățea, iuțea, askuțeam, bănățean, teapă, hoțea, kă, avea, veak, loveam, strivească, vrea, vream, manevrează, auzea, umezeală, zeamă, uzează, ai, aibă, aici, aidoma, kai, tokmai, dai, aflai, grai, haide, mai, maikă, maidan, naiba, nailon, înainte, naiv, arhaizant, niskaiva, spaimă, stai, taină, trai, ia, iar, iarbă, iarnă, iasă, iatak, abia, biată, galaksia, lekția, rodia, diagonală, diagramă, dialog, radiam, iradiat, radiație, terifiant, fiară, likefiat, solia, liant, biliar, miliard, premiat, amiază, ambient, ambientă, astenia, maniak, miniat, funkția, mondial, pensia, diferenția, diferențial, diferențiat, diferențiază, apropiă, pian, piardă, apropiat, piatră, piață, aria, serial, aerian, ilariant, periase, uriaș, speriat, variație, variază,

dreaptă, kafea, fleak, pofta, reglează, răstignească, agreabil, greață, duhnea, prășea, slușească, rușează, kalea, leafa, leagă, aleasă, pilească, ațeleași, mulează, albească, kulmea, întilneam, întilnească, mea, near-

Grup vocalic	Repartiție la nr. de contexte	Grup vocalic	Repartiție la nr. de contexte
ae	4	ea	141
ai	23	ia	79
ao	1	oa	92
au	34	ua	20
ău	13	uă	4
ei	33	ie	75
eo	9	oe	3
eu	19	ue	5
io	31	oi	26
iu	39	ui	64
iu	3	iû	3
ou	8	uo	3

gă, primească, domnea, dumneavoastră, semnează, umplea, puneă, neagră, černeală, spuneam, kimpenească, čineaști, alineat, dimineața, planează, pîndea, gîndească, îndreaptă, 3onglează, aranșează, înseamnă, neant, kontează, mărunțea, desființează, întreagă, rupea, lipsească, pompează, pleacă, prea, preașmă, popreală, lipsea, proptea, rea, real, reakție, akream, podgorean, mireasă, kučerească, fereastră, șireată, virează, sorbea, vorbească, kîrmea, dormeam, urnea, kîrpea, eksersează, kîrtea, dezmořtească, servea, kosea, oboseală, țesem, seară, searbăd, sosească, visează, stea, steag, păstrează, tușea, nădușeală, kunoștea, așteaptă, putea, beteală, teamă, rotească, teatru, viteaz, treabă, treakă, treaz, kățea, iuțeală, askuțeam, bănățean, teapă, hoțeașcă, avea, veak, loveam, strivească, vrea, vream, manevrează, auzea, umezeală, zeamă, uzează, ai, aibă, aici, aidoma, kai, tokmai, dai, aflai, grai, haide, mai, maikă, maidan, naiba, nailon, înainte, naiv, arhaizant, niskaiva, spaimă, stai, taină, trai, ia, iar, iarbă, iarnă, iasă, iatak, abia, biată, galaksia, lekția, rodia, diagonală, diagramă, dialog, radiam, iradiat, radiație, terifiant, fiară, likefiat, solia, liant, biliar, miliard, premiat, amiază, ambient, ambientă, astenia, maniak, miniat, funkția, mondial, pensia, diferenția, diferențial, diferențiat, diferențiază, apropiă, pian, piardă, apropiat, piatră, piață, aria, serial, aerian, ilariant, periase, uriaș, speriat, variație, variază,

scepțional, periodik, superior, sperios, seriozitate, kardiolog, kirlionț, lornion, pasionant, sfios, triodă, staționa, ambițios, avion, viori, aprovizionă, oi, koinčide, kroit, doi, doilea, doină, foi, egoism, 3oi, moi, noi, noile, îndoi, îndoit, apoi, ploi, eroi, eroik, eroițe, eroism, toi, voi, voință, voise, voit, război, iubi, iubit, iubiți, akțiune, rakiu, radiu, sodiul, podium, fiu, fiul, doliu, doliul, domeniu, domeniul, funkțiune, tensiune, opiu, piuliță, propriu, opțiune, ivoriu, kriteriul, kontorsiune, tîrziu, permisiune, skriu, pustiu, kestione, știu, știut, spațiu, spațiul, stațiune, viu, aluviune, viziune, uimire, uita, uiți, trebui, trebuință, trebuit, kuib, lokuind, lokuință, peskuise, peskuit, kuiva, dădui, îngăduință, făgăduit, prăfuișe, fluid, fluizi, prile3ui, prile3uind, prile3uit, nădă3dui, nădă3duit, lui, uluise, keltui, dăinui, kinuise, kinuit, konstruire, instruit, înlănțui, înlănțuire, înlănțuit, pui, îkipuim, îkipuire, îkipuit, epuiza, făptuit, kărărui, ruină, biruință, biruise, biruit, arkuir, verzui, kreskui, skuipa, spui, ačestui, pistrui, obișnui, obișnuit, obișnuiți, alkătui, alkătuit, alkătuiți, prețui, țuikă, prețuind, prețuit, vui, kăzui, brîul, riul, riuri, efektuind, luind, situind, tablou, tablouri, ekoul, kadou, nou, noul, erou, eroul, onktuos, duo, fastuos/ (aer, aevea, maestru, vaer, ea, bea, pribeag, beam, leat, breaz, sticlea, tocmală, acrea, creadă, creare, creată, creastă, creață, lucrează, fixează, pictază, vedea, deal, vedeam, deasă, pîndească, dreaptă, cafea, fleac, poștea, reglează, răstignească, agreabil, greață, duhnea, prăjea, slujească, rujează, calea, leafă, leagă, aleasă, pilească, aceleași, mulează, albească, culmea, întîlneam, întîlnească, mea, meargă, primească, domnea, dumneavoastră, semnează, umplea, punea, neagră, cernelă, spuneam, cîmpenească, cineasți, alineat, dimineața, planează, pîndea, gîndească, îndreaptă, jonglează, aranjează, înseamnă, neant, contează, mărunțea, desființază, întreagă, rupea, lipească, pompează, pleacă, prea, preajmă, popreală, lipsea, proptea, rea, real, reacție, acream, podgorean, mireasă, cucerească, fereastră, șireată, virează, sorbea, vorbească, cîrmea, dormeam, urnea, cîrpea, exersează, cîrtea, dezmorțescă, servea, cosea, oboseală, țeseam, seară, searbăd, sosească, visează, stea, steag, păstrează, tușea, nădușeală, cunoștea, așteaptă, putea, beteală, teamă, rotească, teatru, viteaz, treabă, treacă, treaz, căfea, iușeală, ascuțeam, bănățean, țeapă, hoțescă, avea, veac, loveam, strivească, vrea, vream, manevrează, auzea, umezeală, zeamă, uzează, ai, aibă, aici, aidoma, cai, tocmăi, dai, aflai, grai, haide, mai, maică, maidan, naiba, nailon, înaintea, naiv, arhai-zant, niscaiva, spaimă, stai, taină, trai, ia, iar, iarbă, iarnă, iasă, iatac, abia, biată, galaxia, lecția, rodia, diagonală, diagramă, dialog, radiam, iradiat, radiație, terifiant, fiară, lichefiat, solia, liant, biliar,

miliard, premiat, amiază, ambiant, ambianță, astenia, maniac, mîniat, funcția, mondial, pensia, diferenția, diferențial, diferențiat, diferențiază, apropiat, pian, piardă, apropiat, piatră, piață, aria, serial, aerian, ilariant, periase, uriaș, speriat, variație, variază, urgia, hîrtia, întîrziase, întîrziat, siamez, scria, bestial, moșia, roșiatic, știa, știam, patriarh, beția, terțiar, fluvial, aviator, aviație, nerozie, ziar, dezmiardă, adaos, oală, oameni, oare, oarbă, oase, oaspete, boală, boare, obloane, coace, coadă, coajă, coamă, cucoană, coapsă, coase, coastă, coate, protectoare, cioară, podoabe, doagă, doamnă, doare, doarme, foame, sifoane, foarte, floare, goală, gloată, groapă, groasă, groază, joacă, joace, joasă, mijloace, saloane, paloare, foloase, moale, omoare, frumoase, moarte, noapte, onoare, luminoasă, cunoască, noaten, noastre, cunoaște, bondoacă, întoarcă, întoarsă, întoarce, vapoare, poartă, poate, aproape, proaspăt, proastă, lăptoasă, roade, coroane, răcoroase, roată, cărnoase, persoane, nervoasă, soacră, soare, scoate, pistoale, stoarce, țestoase, castroane, șoptă, șoarece, școală, toamnă, toane, răbdătoare, toarnă, toată, patroană, convoacă, provoace, zăvoare, voastre, zboare, izvoare; au, auxiliar, aud, august, aur, autentic, auzi, cauciuc, caut, precauție, cauză, duceau, dau, adaug, aflau, adăugau, laur, adunare, naufragiu, nautic, mîncau, plîngeau, aranjau, cîntau, pauză, aplaudă, erau, urlau, purtau, sau, scaun, stau, restaurant, arătau; ecuație, efectua, actual, efectuat, dualism, lua, luase, luat, luați, continua, anual, nuanțe, continuare, continuat, continuați, eventual, suav, trotuar, situație, cazul; bău, băut, căuta, adăuga, făuri, lăuda, înăuntrul, nătarău, răul, arnăuți, său, tău, zău; luă, luăm, continuă, continuăm; ei, clei, nucleic, nucleilor, cei, ceilalți, idei, grei, lei, albei, altei, mei, fmeile, doamnei, unei, scînteii, dorinței, apei, cărei, urmei, cîrpei, mersci, artei, fusei, sleiți, stei, tei, proteine, trei, fetitei, vei, vrei, cauzei; ied, ieftin, iepure, ierburi, iēs, ieșire, robie, obiect, biet, client, galaxie, reacție, hărnicie, comedie, audiență, adiere, diez, pudrieră, fier, ficare, fierb, fiert, geologie, felie, relief, palier, palmier, mie, miel, miere, mîez, decembrie, domnie, funcție, pretenție, evidențiere, conștient, entropie, piedică, piele, incipient, pieri, pierd, piesă, piept, pietre, proprie, proprietate, concepție, florărie, orient, experiență, curier, varietate, varieze, urgie, fermier, excursie, hîrtie, țîrzie, osie, scrie, industrie, moșie, știe, apatie, cocotrie, ție, sașietate, vie, viermi, vieți, explozie, dezmiarda; deochi, deodată, deosebit, leopard, uneori, preocupare, preot, teorie, vreo; coeficient, coerent, poet; eu, nucleu, nucleul, liceu, greu, greutate, leu, meu, împreună, neurologic, cupeul, pneumatic, mereu, reumatic, reuși, pa-

teuri, vreun, muzeu, zeul; duel, siluetă, continue, influență, șuetă; ionic, biologic, acționa, radio, mediocru, diodă, fior, primejdios, bu-lion, mînios, funcționa, înfiora, intenționa, pretențioși, pios, excep-tional, periodic, superior, sperios, seriozitate, cardiolog, cîrlionț, lor-nion, pasionant, sfios, triodă, staționa, ambițios, avion, viori, apro-viziona; oi, coincide, croit, doi, doilea, doină, foi, egoism, joi, moi, noi, noile, îndoi, îndoit, apoi, ploi, eroi, eroic, eroice, eroism, toi, voi, voință, voise, voit, război; iubi, iubit, iubiți, acțiune, rachiu, radiu, sodiul, podium, fiu, fiul, doliu, doliul, domeniu, domeniul, funcțiune, tensiune, opiu, piuliță, propriu, opțiune, ivoriu, criteriul, contorsiune, tîrziu, permisiune, scriu, pustiu, chestiune, știu, știut, spațiu, spațiul, stațiune, viu, aluviune, viziune; uimire, uita, uiți, trebui, trebuință, trebuit, cuib, locuind, locuință, pescuise, pescuit, cuiva, dădai, îngă-duință, făgăduit, prăfui, fluid, fluizi, prilejui, prilejuind, prilejuit, nădăjdai, nădăjdut, lui, uluise, cheltui, dăinui, chinuise, chinuit, construi, instruit, înlănțui, înlănțuire, înlănțuit, pui, închipuim, închipuire, închipuit, epuiza, făptuit, cărărui, ruină, biruință, biruise, biruit, arcuit, verzui, creștui, scuipa, spui, acestui, pistrui, obișnui, obișnuit, obișnuieți, alcătui, alcătuit, alcătuiți, prețui, țuică, prețuind, prețuit, cui, căzui; brîul, rîul, rîuri; efectuînd, luînd, situînd; tablou, tablouri, ecoul, cadou, ncu, noul, erou, eroul, onctuos, duo, fastuos).

Pe baza acestor grupări, date de distribuția grupurilor vocalice simetrice de două elemente, vom stabili tipurile de simetrie specifice limbii române, în situația: consoană + vocală + vocală + consoană.

Pentru a sistematiza prezentarea materialului studiat și pentru a facilita studiul comparativ al grupărilor formate după acest model (consoană vocală vocală consoană), am introdus datele obținute în matrice $\frac{V_1 V_2}{V_2 V_1}^{24}$, în care liniile și coloanele notează pe C_1 , respectiv C_2 .

Matricele sînt de tipul graficului de la p. 53.

Pentru fiecare poziție se introduc procentele de apariție a voca-lelor în contextele C_1-C_2 (acestea reprezentînd linia, respectiv coloana, la a căror intersecție se obține poziția studiată). În vederea concentrării datelor, s-a folosit una și aceeași matrice pentru gru-purile vocalice simetrice $V_1 V_2 / V_2 V_1$, înscriindu-se în fiecare poziție procentele corespunzătoare celor două forme (deasupra și dedesub-tul diagonalei poziției respective).

²⁴ V_1 respectiv V_2 notează vocalele, iar C_1 respectiv C_2 notează consoanele.

	p	b	m	f	v	t	ɛ	n	ʃ	s	z	č	g	ʒ	k	g	h	l	r	#
p																				
b																				
m																				
f																				
v																				
t																				
ɛ																				
n																				
ʃ																				
s																				
z																				
č																				
g																				
ʒ																				
3																				
k																				
g																				
h																				
l																				
r																				
#																				

- Diagonala principală a matricei
 = Diagonala poziției respective

Considerînd grupurile vocalice V_1V_2 , respectiv V_2V_1 , în contextele C_1 (anterior grupului), C_2 (posterior grupului), rezultă următoarele cazuri :

$$[1] \quad \begin{cases} C_1V_1V_2C_2 \\ C_1V_2V_1C_2 \end{cases}$$

În acest caz comparăm grupurile simetrice în aceleași contexte (directe), prin citirea pozițiilor din matrice, existente dedesubtul și deasupra diagonalei poziției respective.

$$[2] \quad \begin{cases} C_1V_1V_2C_2 \\ C_2V_1V_2C_1 \end{cases}$$

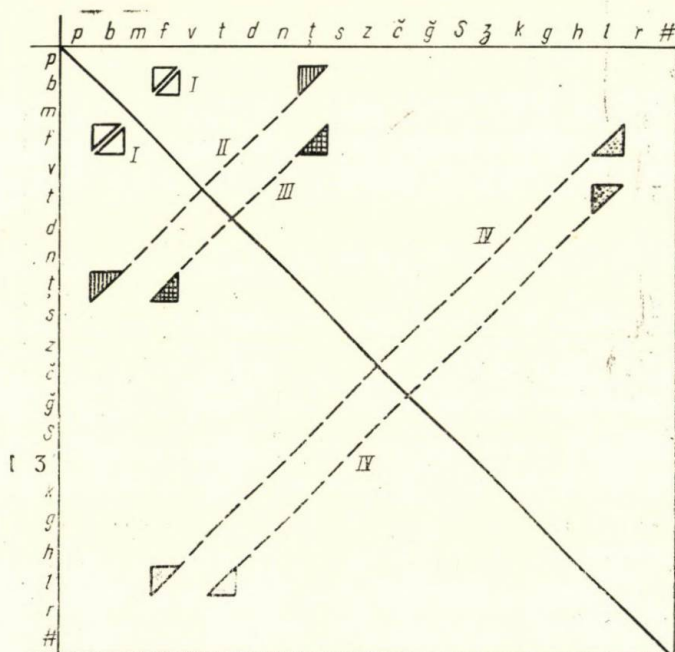
În acest caz comparăm pozițiile pentru același grup V_1V_2 în context direct C_1-C_2 față de context invers C_2-C_1 , prin citirea poziții-

lor din matrice corespunzătoare grupurilor V_1V_2 , simetrice față de diagonală principală a matricei.

$$[3] \quad \begin{cases} C_1V_2V_1C_2 \\ C_2V_2V_1C_1 \end{cases}$$

Comparăm aceleași poziții (ca în cazul [2]), dar pentru grupurile V_2C_1 .

Cazurile :



$$[4] \quad \begin{cases} C_1V_1V_2C_2 \\ C_2V_2V_1C_1 \end{cases}$$

În acest caz comparăm pozițiile pentru același grup V_1V_2 în context direct C_1C_2 față de cele ale grupului V_2V_1 în context invers C_2-C_1 , prin citirea pozițiilor din matrice corespunzătoare grupului V_1V_2 și a simetricilor față de diagonală matricei pentru V_2V_1 . Este o simetrie propriu-zisă față de axa de simetrie a grupului (CV/VC).

$$[5] \begin{cases} C_1 V_2 V_1 C_2 \\ C_2 V_1 V_2 C_1 \end{cases}$$

Coincide cu cazul anterior, [4], deoarece se studiază toate contextele vocalice posibile pentru același grup.

$$[6] \begin{cases} C_2 V_1 V_2 C_1 \\ C_2 V_2 V_1 C_1 \end{cases}$$

Coincide cu cazul [1].

În matrice, cele 4 cazuri se obțin, respectiv, după cum se arată în figura din p. 54.

Pentru grupurile vocalice inventariate de noi, înregistrăm următoarele cazuri de simetrii, respectiv în matricele :

I	<i>ae/ea</i> :	[3]	Cu prob. diferite :	leas (0,018)/seal (0,2)
				leaz (0,07)/zeal (0,05)
				neat (0,11)/țean (0,05)
				reas (0,07)/sear (0,3) ²⁵
II	<i>ai/ia</i> :	[4]	„ „ „	#aer (0,6)/rea# (5,4)
				#aev (0,018)/vea# (0,7)
				#ai# (1,2)/#ia# (0,1)
				#aid (0,018)/dai# (0,05)
		[2]	„ „ „	#iar (0,50)/ria# (0,3)
				#ias (0,07)/sia# (0,16)
		[3]	„ „ „	liar (0,05)/rial (0,1)
				riaz, (0,05)/ziar (0,16)
				#aib (0,24)/bia# (0,28)
				identice# aid (0,018)/dia# (0,018)
III	<i>ao/oa</i> :	[3]	„ „ „	koa3 (0,018)/3oak (0,05)
				doar (0,22)/road (0,018)
				noas (0,05)/soan (0,018)
				roas (0,09)/soar (0,3)
		[4]	„ „ „	roat (0,03)/toar (1,1)
				noat (0,018)/toan (0,018)
		[3]	„ „ „	diferite : kaut (0,07)/kuaț (0,018)
				nau# (0,09)/nua# (0,07)
				naut (0,12)/nuat (0,03)
				#aug(0,05)/gau# (0,018)
IV	<i>au/ua</i> :	[2]	„ „ „	#aur(0,09)/rau# (0,4)
				#aut (0,5)/tau# (0,11)

²⁵ Procentele caracterizează segmentul dat în contextul considerat.

V	<i>ei/ie</i> :	[3]	Cu prob. diferite :	#ied	(0,018)/die#	(0,03)
				#iep	(0,018)/pie#	(0,03)
				#ies	(0,03)/sie#	(0,16)
				#ieş	(0,33)/şie#	(0,07)
				nţier	(0,018)/rientţ	(0,05)
VI	<i>io/oi</i> :	[2]	„ „ „	nios	(0,05)/sion	(0,03)
VII	<i>iu/ui</i> :	[2]	„ „ „	#iuţ	(0,05)/ţiu#	(0,14)
		[3]	„ „ „	#uit	(0,51)/tui#	(0,14)
				ruit	(0,07)/tuir	(0,018)

Celelalte grupuri vocalice — *ău/uă*, *eo/oe*, *eu/ue*, *ou/ua*, *îu/uî* — nu apar niciodată în acelaşi tip de contexte.

Grup de vocale	Tip [1]		Tip [2]		Tip [3]		Tip [4]	
	%		%		%		%	
	ident.	dif.	ident.	dif.	ident.	dif.	ident.	dif.
ae/ea						l-s l-z n-ţ r-s		#-r #-v
ai/ia		#-#		#-d		#-r #-s l-r r-z	#-d	#-b
ao/oa					n-t	k-3 d-r n-s r-s r-t		
au/ua		k-ţ n-# n-t		#-g #-r #-t				
ei/ie						#-d #-p #-s #-ş nţ-r		
io/oi				n-s				
iu/ui				#-ţ		#-t r-t		

Acestea sînt simetrice numai prin structura secvențelor vocale, iar contextele lor nu oferă situații comparabile (cel puțin în materialul parcurs de noi).

Pe baza datelor de mai sus am întocmit un tabel (p. 56) care sintetizează frecvența diverselor tipuri de simetrie în contextele date.

1.3.3. O categorie aparte a grupurilor vocale o constituie secvențele cu ambele elemente identice : $V_1 \equiv V_2$.

Grupurile de acest tip, înregistrate de noi, sînt următoarele : *ee, ii, oo, uu*.

Ele reprezintă 2,3% din totalul grupurilor vocale binare.

Frecvențele care caracterizează fiecare dintre grupurile de mai sus sînt înscrise în tabelul următor. Adăugăm cifrele care notează numărul de contexte diferite în care se distribuie fiecare grup vocalic.

Se observă că grupurile cele mai frecvente sînt cele de tip *ii* (85,1), și cel mai puțin frecvente *uu* (1,5).

Lista completă a contextelor consonantice în care se repartizează aceste grupuri vocale este următoarea :

ee : kl-#, č-l, d-#, f-r, l-(#, l), n-#

ii : b-l, k-ł-l, k-l, č-l, d-l, f-(nd, nł), ğ-l, l-l, m-l, n-l, nf-nł, nłt-nł, nł-l, p-(#, l), pł-l, r-(l, nd), rz-nd, sf-(#, nł), st-l, łt-(nd, nł), ł-l, v-(l, nł, t)

oo : k-rđ, lk-l

uu : n-#

Exemple : /nuklee, rozačeele, orhidee, feerie, alee, releee, graminee; robiile, reacțiile, rokiile, kapričiile, komediile, fiind, ființă, omagiile, familiile, miile, mătăniile, înființa, înștiința, intențiile, kopii, copiilor, konceptiile, teoriile, variind, întîrziind, sfii, sfiičune, desființa, prostiile, știind, știința, relațiile, viilor, kuviință, viitor; koordonare, alkool, kontinuu/ (nuclee, rozaceele, orhidee, feerie, alee, releee, graminee; robiile, reacțiile, rochiile, capriciile, comediile, fiind, ființă, omagiile, familiile, miile, mătăniile, înființa, înștiința, intențiile, copiii, copiilor, conceptiile,

Grup vocalic	%
ee	10,1
ii	85,1
oo	3,2
uu	1,5

Grup vocalic	Repartiție (la nr. de contexte)
ee	7
ii	30
oo	2
uu	1

teoriile, variind, întîrziind, sfii, sfiiciune, desființa, prostiile, știind, știința, relațiile, viilor, cuviință, viitor; coordonare, alcool; continuu).

Singurul caz de simetrie pe care îl putem semnaliza este acela al contextelor consonantice, și anume al contextelor particulare $C_1 \equiv C_2$.

ee : 1-1 (2,8)

ii : 1-1 (2,8)

1.3.4. Grupurile vocalice asimetrice sînt grupurile consonantice care prezintă o singură orientare posibilă, prin permutarea lor nemai-rezultînd grupuri proprii limbii române.

Citirea lor este posibilă într-un singur sens, considerînd un ventil [$\rightarrow$], care indică orientarea unică :

$$C_1V_1 \rightarrow V_2C_2$$

$$C_2V_1 \rightarrow V_2C_1$$

Dintre grupurile vocalice liniare inventariate de noi, numai 3,1% sînt grupuri asimetrice.

Acestea sînt : *ăi*, *îi*.

Grup vocalic	%
ăi	27,05
îi	72,9

Grup vocalic	Repartiție (la nr. de contexte)
ăi	20
îi	11

Frecvența și posibilitatea lor de distribuire în diverse contexte sînt :

Observăm că grupul *îi* are o probabilitate de apariție mai mare (în timp ce *ăi* se distribuie într-un număr mai larg de contexte). Lista completă a contextelor consonantice în care se distribuie grupurile vocalice asimetrice cu două elemente este următoarea : *ăi* : k-(#, n, t), d-(#, l, Ț), g-n, h-Ț, m-k, p-#, s-#, sp-m, str-n, t-(#, Ț), tr-(m, nd, t, Ț), v-t.

îi : #-#, b-g, k-(n, r, t), f-#, m-(#, n), nt-(#, l), p-n.

Exemple : /orăkăi, kăina, pokăit, odăi, odăile, odăiță, găină, buhăiți, măikuță, pipăi, săi, înspăimînta, străin, tăi, tăiș, trăim, trăind, trăit, trăiți, văita; îi, bîigui, kîine, rikîiră, sîkîit, gîfii, rămîi, mîine, întîi, întîile, pîine/ (orăcăi, căina, pocăit, odăi, odăile, odăiță, găină, buhăiți, măicuță, pipăi, săi, înspăimînta, străin, tăi, tăiș, trăim, trăind, trăit, trăiți, văita; îi, bîigui, cîine, rîcîiră, sîcîit, gîfîi, rămîi, mîine, întîi, întîile, pîine).

Data fiind structura acestor grupuri, care nu permit permutarea elementelor lor, discutarea aspectului simetric este posibilă numai pe baza contextelor consonantice.

Introducînd datele obținute în matrice V_1V_2 , ca și în cazul grupurilor vocalice simetrice, singurele contexte simetrice care rezultă din cercetarea grupurilor *ăi*, *îi* sînt $\#-\#$ pentru *îi* (secvență fonematică despre care trebuie, de altfel, să specificăm că reprezintă în același timp un cuvînt independent).

1.3.5. Cele mai largi segmente complexe vocalice românești sînt cele de *t r e i* și *p a t r u* vocale, de tipurile: $V_1V_2V_3$; $V_1V_2V_3V_4$.

Din materialul excerptat de noi au rezultat următoarele grupuri vocalice cu structura $V_1V_2V_3$:

aia, aio, oio, aiu, ăia, eai, eau, eia, eiu, eoa, eoi, iau,iei, ioa, iua, îia, oai, oia, oio, oiu, oua, ouă, uai, uau, uia, uio, uoa.

Analizînd aceste grupuri din punct de vedere al structurii lor vocalice, notăm — de la stînga — vocalele, în raport cu locul (întîi, al doilea și al treilea) pe care îl ocupă în grup și obținem următoarea situație:

- locul I: a, ă, e, i, î, o, u;
- locul II: a, e, i, o, u;
- locul III: a, ă, i, o, u

Locurile I, II, III se definesc, respectiv, astfel: spunem că o vocală ocupă locul I atunci cînd într-o secvență de trei poate fi urmată de altele două, iar imediat după ea urmează măcar una dintre vocalele de pe locul II; spunem că o vocală ocupă locul II atunci cînd într-o secvență de trei poate fi precedată măcar de una dintre vocalele de pe locul I și urmată măcar de una dintre vocalele de pe locul III; spunem că o vocală ocupă locul III atunci cînd într-o secvență de trei este precedată de două vocale și urmează măcar unei vocale de pe locul II ²⁶.

Rezultă că:

- Primul loc poate fi ocupat de toate vocalele.
- De pe locul al doilea sînt excluse vocalele centrale *ă*, *î*, toate celelalte cinci vocale putînd apărea în această poziție.

²⁶ Spre deosebire de grupurile binare, care se definesc prin însăși structura lor simetrică sau asimetrică, în cadrul grupurilor ternare considerăm necesară o descriere a componenței, dată fiind complexitatea lor, care introduce mai multe elemente condiționate.

— Pe locul al treilea pot apărea de asemenea cinci vocale, cele excluse fiind în acest caz vocala anterioară *e* și vocala centrală *î*. Pe fiecare dintre cele trei locuri pot apărea toate vocalele posterioare, vocala anterioară *i* și vocala centrală *ă*.

Oricare vocală poate apărea cel puțin într-o poziție, în grupurile de trei vocale.

În tabelele care urmează înregistrăm frecvența acestor grupuri și distribuția lor (numărul de contexte pentru fiecare grup) :

Grup vocalic	%
aia	4,9
aio	0,28
aiu	1,3
ăia	4,9
eai	1,3
eau	25,5
eia	9
eiu	0,6
eoă	1,05
eoî	0,6
iai	0,6
iau	1,3
iei	1,6
ioa	14,8
iua	2,5
îia	1,3
oai	0,28
oia	4
oio	0,8
oiu	2,1
oua	2,8
uai	0,28
uau	0,28
uia	14
uio	0,6

Din acest tabel reiese că cele mai frecvente grupuri vocalice cu trei elemente sînt : *eau*, *ioa*, *uia*.

Lista completă a contextelor consonantice în care se distribuie grupurile vocalice cu trei elemente este următoarea ²⁷ :

aia : #-#, d-#, t-#
aio : m-n, ; *aiu* ; #-r, *ăia* : b-t, f-#, p-n3, t-(#, s, t), tr-(#, sk) ; *eai* : mț-#, t-# ; *eau* : b-#, kl-#, d-#, gn-#, l-#, ln-#, m-#, mpl-#, n-#, n-r, nd-#, p-#, ps-#, r-#, s-#, {n-#, t-#, d-n, v-#, vr-#, z-# ; *eia* : l-#, m-# ; r-#, rd-t, st-#, tr-# ; *eiu* : ĉ-l, ț-l ; *eoă* : pl-p ; *eoî* : gr-# ; *iai* : {t-# ; *iau* : #-#, {t-# ; *iei* : #-#, bl-#, f-#, ģ-#, l-#, m-#, n-#, nț-#, r-#, s-#, str-#, t-#, ț-#, z-# ; *ioa* : bț-r, d-s, 3d-s, l-n, mn-s, n-s, nd-s, nf-r, r-(d, r, s), ț-s, v-(n, r) ; *iua* : z-# ; *îia* : nt-# ; *oai* : z-k ; *oia* : lt-#, nd-(l, m), p-n, v(#, m), vr-# ; *oie* : b-r, nd-(l, sk), pr-kt, v-# ; *oio* : nd-s, v-s, v-ș ; *oiu* : s-l, zb-l ; *oua* : d-#, n-#, pl-# ; *ouă* : #-(#, l), d-#, n-#, nd-#, pl-# ;

uai : l-# ; *uau* : kt-# ; *uia* : b-#, bz-#, k-(#, sk), d-(#, l), g-#, l-#, lt-l, m-t, n-(#, l), p-#, pt-#, r-#, rg-#, rp-#, s-(#, l), st-#, {n-#, t-# ; *uio* : d-ș, nd-ș.

²⁷ Menționăm că în această listă am introdus numai exemplele înregistrate de noi în textele parcurse și am făcut abstracție de o serie de alte contexte cunoscute în limbă. De exemplu : *aio* : b-n (*baionetă*), *oai* : s-k (*săsoaică*), *oio* : -s(*soios*), *oiu* : r-l (*roiul*), *oua* : pl-t (*plouat*) etc.

Exemple : /aia, odaia, bătaia ; maioneză ; aiurit ; băiat, lăfăia, paianjen, tăia, tăiase, tăiat, trăia, trăiască, simțea, puteai ; beau, sticleau, dădeau, răstigneau, okoleau, întâlneau, temeau, umpleau, veneau, pîndeau, rupeau, vopseau, mureau, găseau, posteau, foșneau, roteau, totdeauna, aveau, vreau, auzeau, aceleia, femeia, reia, ardeiat, acesteia, treia, obișniul, țiteiul ; pleoape ; greoi ; știa ; iau, știau ; ie, bibliei, filozofiei, tehnologiei, familiei, economiei, telefoniei, garanției, cofetăriei, expresiei, măiestriei, cutiei, soției, invaziei, subțioară, radioasă, primejdioasă, medalioane, calomnioase, mînioase, grandioasă, înfioară, perioadă, superioară, sperioasă, prețioasă, avioane, vioară ; ziua ; întâia ; franțuzaică ; altoia, îndoială, îndoiam, poiană, voia, înkovoiam, vroia ; îndioș, voios, voioșie ; soiul, războiul, doua, noua, ploua, luai, efektiv, trebuia, kibzuia, lokuia, lokuiașcă, făgăduia, făgăduială, bîguia, kibzuia, ațelua, keltuială, muiat, nuia, bănuială, înkipuia, făptuia, kăruia, tîrguia, șerpuia, suia, îngeșuială, ațestuia, obișnuia, statuia ; duioșie, înduioșă/. (aia, odaia, bătaia, maioneză ; aiurit ; băiat, lăfăia, păianjen, tăia, tăiase, tăiat, trăia, trăiască, simțea, puteai, beau, sticleau, dădeau, răstigneau, ocoleau, întâlneau, temeau, umpleau, veneau, pîndeau, rupeau, vopseau, mureau, găseau, posteau, foșneau, roteau, totdeauna, aveau, vreau, auzeau, aceleia, femeia, reia, ardeiat, acesteia, treia ; obișniul, țiteiul ; pleoape ; greoi ; știa ; iau, știau ; ie, bibliei, filozofiei, tehnologiei, familiei, economiei, telefoniei, garanției, cofetăriei, expresiei, măiestriei, cutiei, soției, invaziei ; subțioară, radioasă, primejdioasă, medalioane, calomnioase, mînioase, grandioasă, înfioară, perioadă, superioară, sperioasă, prețioasă, avioane, vioară ; ziua ; întâia ;

Grup vocalic	Repartiția (la nr. de contexte)
aia	3
aio	1
aiu	1
ăia	8
eai	2
eau	22
eia	6
eiu	2
eoă	1
eoî	1
iaî	1
iau	2
iei	14
ioa	14
iua	1
îia	1
oai	1
oia	7
oio	3
oiu	2
oua	3
ouă	6
uai	1
uau	1
uiu	22
uio	2

franțuzoică; altoia, îndoială, îndoiam, poiană, voia, încovoiam, vroia; îndoiș, voioș, voioșie; soiul, războiul, doua, noua, ploua; luai; efectuau, trebuia, chibzuia, locuia, locuiască, făgăduia, făgăduiala, biiguia, chibzuia, aeluia, cheltuială, muiat, nuia, bănuială, închi-puia, făptuia, căruia, târguia, șerpuia, suia, înghesuială, acestuia, obișnuia, statuia; duioșie, înduioșă).

Din examinarea listei contextelor consonantice rezultă că:

a) în poziție inițială apar: *aiu*

b) celelalte segmente apar în poziție medială și (mai ales) în poziție finală.

Următoarele segmente de trei vocale pot alcătui cuvinte independente: *aia*, *iau*²⁸.

Materialul analizat poate fi introdus în matrice de tipul $V_1V_2V_3$, în care liniile și coloanele reprezintă contextele consonantice C_1 și C_2 ale grupurilor vocale.

Tipurile de simetrie se stabilesc în asemenea cazuri pe baza contextelor consonantice. Ne referim la posibilitatea permutării elementelor consonantice, grupurile vocale rămânând invariabile.

Pentru această situație am inventariat numai ocurența grupurilor vocale în contextele simetrice, de o formă particulară, și anume:

$$C_1 \equiv C_3$$

Matrice: *aia*: #-#(2,1)

ăia: t-t (0,8)

iau: #-#(1,05)

iei: #-#(0,28)

Interpretînd elementul medial al acestor secvențe ca „axă de simetrie”, putem degaja următoarele structuri simetrice: *aia*, *iei*, *oio*, *uau*. Vocalele extreme se află într-o poziție de „răsturnare” simetrică față de axe de simetrie ale secvențelor, în cazurile: *aio/oia*, *aiu/uio*, *iau/uai*, *ioa/oai*, *oiu/uio*.

Din materialul parcurs de noi au rezultat următoarele grupuri de patru vocale, de tip VVVV²⁹: *ăiau*, *eaia*, *eaei*, *eiau*, *eoae*, *ioae*, *îiau*, *îioa*, *oiai*, *oioa*, *uiau*.

²⁸ Secvența *ăia*, de exemplu, face parte din aceeași categorie, dar în materialul parcurs de noi nu a apărut în această calitate, ci numai inclusă în alte cuvinte.

²⁹ Dintre grupurile cu mai mult de două vocale, inventariate de noi, circa 95,27% sînt de trei vocale, iar 4,47% de patru vocale. Am grupat laolaltă aceste două categorii de segmente (atunci cînd am calculat frecvențele comparative, și nu atunci cînd am distins calitatea fonemelor), avînd în vedere frecvențele lor caracteristice foarte mici.

Probabilitatea de apariție și distribuția acestora sint următoarele :

Aceste grupuri nu sînt prea frecvente în limba română, nediferind prea mult în același timp, între ele, nici ca frecvență, nici ca distribuție.

Contextele în care se distribuie sînt :
ăiau : tr-# ; *eaua* : n-# ; p-# ; *caei* : mn-# ;
eiau : nt-# ; *eoae* : gr-# ; *ioae* : v-# ; *îiau* :
 rț-# ; z-# ; *îioa* : r-s ; *oiai* : v-# ; *oioa* : v-s ;
uiiau : 3-# , m-# .

Pentru fiecare grup vocalic înregistrăm deci următorul număr de contexte :

Exemple : /trăiau ; kafeneaua, kanapeaua ; dumneaei ; skinte-
 iau ; greoae ; vioae ; skirțiau, bîziiu ; rîioasă ; voiai ; voioasă, stră-
 3uiiau, vămuiiu/ (trăiau ; cafe-
 neaua, canapeaua ; dumneaei ;
 scînteiau ; greoae ; vioae ; scîrțiau,
 bîziiu ; rîioasă ; voiai ; voioasă ;
 străjuiau, vămuiiu).

Nici unul dintre aceste grupuri nu poate apărea în poziție inițială, iar în majoritatea cazurilor apar la finală.

Structura secvențelor : consoană — vocală — vocală — vocală — vocală — consoană — nu ne-a permis, pe baza materialului parcurs, degajarea de contexte simetrice.

Grup vocalic	%
ăiau	0,28
eaua	0,56
caei	0,28
eiau	0,79
eoae	0,28
ioae	0,28
îiau	0,60
îioa	0,28
oiai	0,28
oioa	0,28
uiiau	0,28

Grup vocalic	Repartiție (la nr. de contexte)
ăiau	1
eaua	2
caei	1
eiau	1
eoae	1
ioae	1
îiau	2
îioa	1
oiai	1
oioa	1
uiiau	2

1.4.1. Grupuri consonantice

De astă dată, descrierea noastră pornește de la mulțimea de consoane $C = \{b_1, b_2 \dots b_n\}$ unde b_1, b_2 etc. simbolizează toate consoanele limbii, și de la mulțimea de vocale $V = \{a_1, a_2 \dots a_n\}$ unde a_1, a_2 etc. simbolizează toate vocalele limbii. Am urmărit acum combinațiile posibile între consoanele $b_1, b_2 \dots b_n$, grupate câte două și distribuția lor în contextele vocalice $a_1, a_2 \dots a_n$, la care am adăugat posibilitatea : pauză (#). Prin *poziție distribuțională* vom înțelege locul ocupat

într-un șir de foneme, de consoane, definite în relație cu un context specificat; *contextul* va fi constituit de vocale, care precedă sau urmează grupul în șirul respectiv.

Cercetarea secvențelor formate din consoană + consoană pune în lumină două tipuri de grupuri consonantice. Definim ca *grupuri consonantice simetrice* grupurile în care, permutând cele două elemente între ele, obținem grupuri proprii limbii române.

Am inventariat circa 10.000 de grupuri consonantice cu două elemente, dintre care circa 2/3 sînt grupuri simetrice (indiferent de cele trei poziții posibile: inițială, medială, finală), raportul simetric/asimetric fiind deci altul în cazul consoanelor, decît în cel al vocalelor.

Au rezultat următoarele grupuri consonantice simetrice distincte:

bl/lb, br/rb, bz/zb, kl/lk, kn/nk, kr/rk, ks/sk, kş/şk, dn/nd, dr/rđ, fl/lf, fr/rf, gn/ng, gr/rg, lp/pl, ls/sl, tl/tl, lv/vl, lz/zl, mn/nm, nr/rm, ms/sm, mt/tm, nr/rn, ns/sn, nt/tn, nz/zn, pr/rp, ps/sp, rt/tr, rv/vr ³⁰.

Diferențele de frecvență dintre grupuri se reflectă în tabelul de la p. 65.

Din acest tabel rezultă că grupurile cele mai frecvente sînt: *nd* (10,2) și *nt* (10,2); grupurile cel mai puțin frecvente sînt: *ks, dn, gn, ms, nm, nr, tl, tn* și *vl* (0,02).

Lista completă ³¹ a contextelor vocalice în care se distribuie grupurile consonantice (cu două elemente) simetrice este următoarea:

bl: #-(-ă, e, o, u), a(-a, e, o), ă(-i), i(-i), o(-e, i, o), u(-i);

lb: a(-#, a, ă, e, u), ă(-a, ă, e), o(-#), u(-e, u);

br: #-(-a, ă, i, î, o, u), a(-i), e(-a), o(-a, u, e);

rb: a(-i, o), ă(-ă, i, u), e(-e, i), o(-#, a, ă, e, i, u), u(-a, i);

bz: i(-u);

³⁰ Se știe că în limba română mai există cîteva asemenea grupuri, dar ele n-au apărut în textele excerptate, fapt din care rezultă că au o mai mică probabilitate de apariție. Astfel de grupuri sînt, de exemplu: *gl/lg* (în: *glosar* și *algă*), *hr/rh* (în: *hrană* și *orhidee*), *mz/zm* (în: *zumet* și *zmeu*) etc. În materialul parcurs de noi, am înregistrat grupurile: *rh, gl, zm* etc., dar n-am înregistrat aceleași grupuri în forma lor permutată, care reprezintă condiția simetriei (respectiv: *hr, lg, mz* etc.).

³¹ Ca și în cazul grupurilor consonantice (v. mai sus nota 30), am introdus în această listă numai exemple înregistrate de noi în textele pe care le-am parcurs, făcînd, deci, abstracție de alte contexte cunoscute în limbă. De exemplu: *kl*: e-i (*sfecli*), *ks*: a-e (*axe*), *mb*: a-a (*ambala*), *nv*: a-e (*anvelopă*), *nz*: i-# (*minz*), *tl*: e-u (*pecetlui*), *vl*: #-ă (*vlastar*), *zl*: #-o (*zloată*) etc.

Grup cons.	%	Grup cons.	%
bl	1	lb	0,9
br	2,02	rb	1,2
bz	0,02	zb	0,3
kl	1,05	lk	0,7
kn	0,07	nk	2,04
kr	2,9	rk	0,8
ks	1,2	sk	4,7
kʃ	0,02	ʃk	0,4
dn	0,02	nd	10,2
dr	1,5	rd	1,4
fl	0,7	lf	0,2
fr	1,3	rf	0,4
gn	0,02	ng	1,6
gr	3,4	rg	1,07
lp	0,3	pl	3,05
ls	0,07	sl	0,5
lt	4,6	tl	0,02
lv	0,5	vl	0,02
lz	0,05	zl	0,05
mn	1,9	nm	0,02
mr	0,02	rm	3,05
ms	0,02	sm	0,4
mt	0,1	tm	0,3
nr	0,02	rn	1,1
ns	3,5	sn	0,1
nt	10,2	tn	0,02
nz	0,8	zn	0,1
pr	9,7	rp	0,2
ps	0,2	sp	4,9
rt	4,9	tr	6,5
rv	0,6	vr	1,1

zb : #(-a, o, u), ă(-a, o), i(-e, i) ;

kl : #(-a, ă, i, o), ă(-ă, i), e(-a, e, i, o, u), u(-e) ;

lk : a(-a, ă, î, o, u), ă(-î), a(-u), o(u), u(-a, u) ;

kn : i(-e, i), o(-e)

nk : a(-#), i(-a, o), î(-a, ă, î, o, u), o(-a, u), u(-a, ă) ;

kr : #(-a, ă, e, i, o), a(-ă, i), ă(-i), e(-e, o, u), i(-i), o(-a, o),
u(a, ă, e, u) ;

rk : a(-#, a, ă, u), ă(-a), e-(-#, a, u), i(-a, o, u), î(-u), u(-a) ;

ks : a(-#, a, i), e(-#, a, ă, e), e(-i, o), i(-#, a), o(-i), u(-o) ;

sk : #(-a, ă, î, o, u), a(-#, ă, u), ă(-a, ă, o, u), e(-#, ă, î) ;

kʃ : e(-i) ;

{k : #-(o), a(-a,), i(-ă), o(-a);
dn : e(-i);
nd : a(-#, a, ă, e, i), e(-a, e, i), i(-#, a, e), i(-i, u), î(-#, a, ă, e, i, o, u), o(-#, a, e, i, o, u), u(-#, a, e, u);
dr : #(-a, e, o, u), a(-e, u), i(-i, o), o(-i);
rd : a(-#, e, a, o), e(-e, u), o(-#, a, e, i, o), u(-a);
fl : #(-a, ă, o), a(-a, ă, i), u(-ă, a, e);
lf : o(-#), u(-#, u);
fr : #(-a, ă, e, i, î, u), i(-e), o(-a);
rf : a(-ă, u), e(-e, o), î(-#, u), o(-o);
gn : i(-o);
ng : a(-a), i(-a, ă, î, u), î(-#, a, ă, î, u), u(-#, a, ă, u);
gr : #(-a, ă, e, i, î, o, u), a(-a, ă, e, i, o), e(-a, e, i, u), i(-e), o(-a, e), u(-ă);
rg : a(-#, a, ă), e(-#, a, ă, î), o(-a), u(-#, u);
lp : a(-ă, i), î(-#), u(-e);
pl : #(-a, ă, e, i, î, o), a(-a, i), e(-a, i), i(-ă, i, o), u(-i);
ls : u(-#, a, ă);
sl : #(-a, ă, o, u), î(-e);
lt : a(-#, a, ă, e, i, o, u), ă(-a, ă), e(-e, u), o(-a, ă, e, i), u(-#, a, ă, e, i, î, o, u);
tl : i(-au);
lv : a(-a, e) î-ă, o(-a, ă, e);
vl : #(-a);
lz : ă(-e),
zl : ă(-e), e(-e);
mn : a(-a, ă, e), e(-#, a, ă, e, i, u), i(-i), o(-i, u), u(-e, i);
nm : î(-o);
mr : #(-e);
rm : a(-a, ă, e), ă(-a), e(-a, e, i), i(-#, a, ă, i), î(-u), o(-#, a, ă, e, i), u(-a, ă, e, î);
ms : a(-a);
sm : #-î, i(-#, e, u);
mt : i(-ă, e);
tm : a(-o, i(-#, e, i));
nr : î(-u);
rn : a(-ă, e, i), e(-#, a, e, i, u), i(-ă, i), o(-a, ă, e, i), u(-a, ă, e, i);
ns : a(-a), i, e(-ă, i, u), i(-#, ă, e, u), î(-#, a, ă, e, o, u), o(-a, e, i), u(-#, e);

sn : #(-o), e(-e, i), o(-i);
nt : a(-#, a, e, i, o, u), e(-#, ă, e, i, u), i(-#, a, e, i), î(-#, a, ă, e, i, î, o, u), o(-a, e, i, o, u), u(-a, ă, e, u);
tn : e(-i);
nz : e(-i), î(-a, ă, e, i, î, u), o(-#, u, ă, e, î);
zn : a(-i), ă(-e, î), e(-a);
pr : #(-a, ă, e, i, o), a(-a, e, i, o), ă(-a, u), e(-e, i, o), o(-e, i), u(-a, -i);
rp : a(-a), î(-e), e(-e);
ps : i(-ă, î), o-i;
sp : #(-a, ă, e, i, î, o, u), a(-ă, e), ă(-î, u), e(-ă), e(-e, o, u), i(-ă, i, o, u), o(-o), u(-e, i);
rt : a(-a, ă, e, i, u), ă(-a, ă, u), e(-#, a, e), î(-i), o(-#, a, ă, i, o, u), u(-a, ă, e, î, o);
tr : #(-a, ă, e, i, î, o, u), a(-a, ă, i, u), ă(-a, ă, e, î), e(-a, ă, e, i, o), i(-e, u), o(-i, o);
rv : e(-#, a, ă, e, i), u(-e);
vr : #(-a, e, î), e(-e, o).

Numerele de contexte în care se distribuie fiecare grup consonantic simetric sint cuprinse în tabelul de la p. 68.

Exemple : /blănuri, bleg, blok, bluză, șablon, tăblie, biblie, goblen, oblik, oblon, table, sublim; alb, albastru, albe, salba, album, sălbatek, înălbești, kolb, pulbere, tulbura; braț, brădet, briçag, brîncă, bronșită, broaște, brusk, fabrică, febra, fibra, obraz, sobru, șubred; karbid, arbore, sărbători, bărbie, kărbune, fierbe, fierbinte, crb, čorba, čorbă, čorbe, sorbea, sorbi, orbul, urban, turbină; kibzui; zbate, zboară, zbură, răzbate, război, izbea, izbi; klătină, klăti, klin, klonț, năklăi, făkli, sfekla, stikle, stiklea, stikli, čiklon, čiklu, bukle; kalkar, kalkă, salkim, alkoolik, balkoane, kalkul, kălkind, polkuță, kulka, kulkuș; iknește, pokni, poknește; țank, zinkat, inkomparabil, stîнка, stînkă, înkîntat, stînkos, înkoače, înkurkat, kronkāni, konkurs, lunka, lunkă; kramă, krăpa, krește, krimă, kronkānit, akră, akri, înăkri, sekret, nekrofor, rekrut, sikriu, auto-krat, okroti, lukra, lukră, lukrez, lukrează, lukru; pork, parka, parkă, parkuri, înkărka, înčerck, înčerka, înčerku, čirka, čirkomvoluțiuni, čirkul, sfîrkul, urka; aks, taksă, aksiomă, indeks, eksağera, koneksă, eksekuta, leksik, eksotik, fiks, fiksa, toksik, luksoasă; skaun, skăpa, skîrții, skop, skoală, skump, teask, kreaskă, askut, kăska, năskător, răscoală, păskut, kresk, kreskători, kreskind, des- koperi, deskusut, riska, iskăli, iskodi, iskusit, mosk, kunoskut, uska,

krepuskul; indeksî, şkolar, şkoală, taşka, mişka, mişkă, kloşka; vrednik; doktorand, sanda, banda, kandelă, kandidat, suspendat, endem, apendice, citind, şcinda, şcindez, şcindează, indispus, indus, văzînd, îndată, îndelete, pîndi, îndoi, îndura, rond, sonda, rondel,

Grup cons.	Repartiție (la nr. de contexte)	Grup cons.	Repartiție (la nr. de contexte)
bl	13	lb	11
br	11	rb	15
bz	1	zb	7
kl	12	lk	10
kn	4	nk	12
kr	18	rk	13
ks	13	sk	25
kş	1	şk	4
dn	1	nd	30
dr	8	rd	13
fl	9	lf	3
fr	8	rf	7
gn	1	ng	14
gr	20	rg	10
lp	4	pl	14
ls	3	sl	5
lt	22	tl	1
lv	6	vl	1
lz	1	zl	2
mn	14	nm	1
mr	1	rm	21
ms	1	sm	4
mt	2	tm	4
nr	1	rn	18
ns	20	sn	4
nt	32	tn	1
nz	12	zn	4
pr	18	rp	3
ps	4	sp	22
rt	23	tr	24
rv	7	vr	25

kondiție, bondoacă, fondul, rotund, funda, unde, askundea, fundul; draperie, drept, dreaptă, drob, drum, adresă, kadru, klorhidrik, hidrofor, kodri; ard, arde, gardă, kardiak, pardoseală, verde, pierdut, fiord, mordant, kordeluță, sordid, kordon, surda; flakăra, flămînd, floare, afla, află, afli, sufla, suflă, sufle; golf, sulf, sulfurik; frază,

frăgezime, frenetik, frigă, frînge, frukt, čifre, šofran ; harfă, parfum, ferfeniță, perfora, virf, virful, morfologie ; ignora ; angaža, ăingaș, stingă, stingînd, lingură, plîng, ăingav, plîngă, plîngînd, ăngust, ung, lungă, lungă, lunguiet ; gras, grăbit, greși, greșea, grimasă, grîne, gros, groasă, grumaz, agrar, neagră, agreabil, agrikol, agronom, integra, negre, ănegri, negru, digresiune, ogradă, progresă, zugrăvi ; sparg, argat, argăți, merg, tergal, čergă, mergînd, organik, kurg, murgul ; talpă, alpin, stîlp, vulpea ; plan, plăkut, pleka, pleacă, plin, plîns, plod, ploaie, aplană, aplikă, deplin, țiplă, disăiplină, diplomă, supliment ; muls, pulsație, mulsă ; slab, slănină, slobod, slugă, vîsle ; alt, alta, daltă, alte, saltea, saltimbank, palton, saltul, sălta, săltă, deltei, keltui, bolta, boltă, bolte, boltit, mult, sultan, kultă, adulte, ultima, insultînd, tumultos, vultur ; titlu ; halva, alveolar, vîlvă, rezolva, rezolvă, rezolve ; vlagă ; ănkălzește ; răzleț, dezlega ; amnar, doamnă, doamne, semn, semna, semnă, semne, temniță, lemnul, skimmnik, somnifer, somnul, dumnezeu, dumneavoastră, pumni ; ănmormînta ; mreană ; armaș, armă, parmezan, sărman, termal, ferme, terminal, infirm, infirma, firmă, infirmi, kirmui, enorm, formal, formă, forme, formînd, urmaș, urmă, turme, urmînd ; samsar ; smîntînă, turism, idiotisme, optimismul ; simță, simte ; atmosferă, ritm, aritmetikă, ritmik ; ănrudit ; toarnă, karne, karnea, karnivor, infern, hibernă, perne, vernisaș, černută, bîrnă, stîrni, orna, bornă, borne, časornik, turna, urnă, turnesol, urni ; ansamblu, ekspansiv, densă, densitate, tensiune, pensulă, prins, ăntinsă, prinse, insultă, plîns, ăînsak, plînsă, ănsetat, ănseamnă, ănsorit, ănsoară, ănsura, konsakrat, konsekvent, konsistent, uns, unse ; snop, lesne, ănlesni, šosnik ; neant, panta, pantă, konstante, antinomie, pantof, dantură, strident, inventă, lentă, lente, dentiție, absolventul, ăgînt, ănainte, ănaintează, intim, vînt, kînta, kîntă, kînte, trîntea, ăntins, ăntilni, întokmi, ăntoarče, ăntunerek, kontabil, kontează, kontinent, kontor, kontur, nunta, nuntă, punte, untură ; etnik ; kenzină, ămpinzi, pinza, brînză, pînze, vînzînd, prînzul, bronz, bronzul, frunză, frunze, tunzînd ; praznik, trăznește, trăzni, beznă ; prag, prăda, prepeliță, prea, prim, program, proastă, kapra, kapre, aprig, aprod, aproape, kăprar, kăprui, depresiune, deprimant, reproș, opreliște, opri, supra, kuprins ; harpa, kîrpe, kîrpeală, serpentină ; lipsă, lipsi, oropsi ; spanak, spătar, spečal, spital, spîn, sport, spune, proaspăt, aspekt, răspîndi, răspăr, răspuns, despăgubi, lespede, despotik, nespus, ispăși, ispiti, disponibil, dispus, gospodar, suspenda, suspin ; arta, artă, parte, karteă, artist, kartuș, depărta, depărtă,

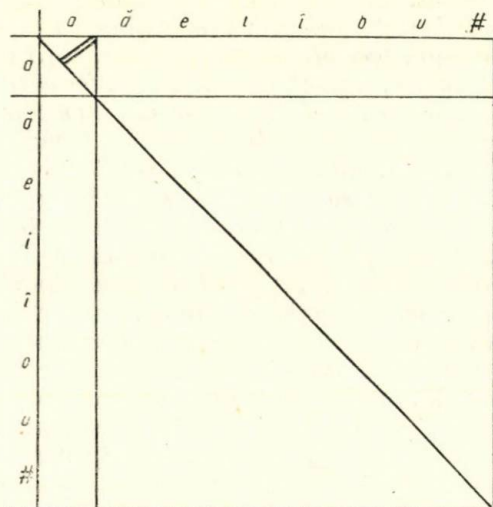
čert, čerta, čerte, hîrti, port, sorta, aortă, portik, ortoped, mortul, purta, purtă, kurtezan, kurtînd, burtos; traĝe, trăda, tresar, trea-kăt, trimite, trînti, tropot, trupă, patra, atrăgînd, patrimoniu, pa-tru, pătrat, pătrăţea, pătruns, către, bătrîn, metra3, kronometra, vetre, metrik, metronom, vitreg, potrivit, şotron; nerv, inerva, vervă, pervers, inervează, servil, survenit; vraf, vreme, vreo, vrînd, manevre, nevropat/ (blănuri, bleg, bloc, bluză, şablon, tăblie, biblie, goblen, oblic, oblon, table, sublim; alb, albastru, albe, salba, album, sălbatec, înălbeşti, colb, pulbere, tulbura; braţ, brădet, briceag, brînci, bronşită, broaşte, brusc, fabrică, febra, fibra, obraz, sobru, subred; carbid, arbore, sărbători, bărbie, cărbune, fierbe, fierbinte, orb, ciorba, ciorbă, ciorbe, sorbea, sorbi, orbul, urban, turbină; chibzui; zbate, zboară, zbura, răzbate, război, izbea, izbi; clătină, clăti, clîn, clonţ, năclăi, făclii, sfecla, sticle, sticlea, sticli, ciclon, ciclu, bucle; calcar, calcă, salcîm, alcoolice, balcoane, calcul, călcînd, polcuţă, culca, culcuş; icneşte, pocni, pocneşte; ţanc, zincat, incomparabil, stîncea, stîncă, încîntat, stîncos, încoace, încurcat, croncăni, concurs, lunca, luncă; cramă, crăpa, creşte, crimă, croncănit, acră, acri, înăcri, secret, necrofor, recrut, sicriu, autocrat, ocroti, lucra, lucră, lucrez, lucrează, lucru; porc, parca, pareă, parcuri, încărcă, încerc, încerca, încercui, circa, circumvoluţiuni, circul, sfîrcul, urca; ax, taxă, axiomă, index, exagera, conexă, executa, lexic, exotic, fix, fixa, toxic, luxoasă; scaun, scăpa, scîrţii, scop, scoală, scump, teasc, crească, ascut, căsca, năs-cător, răscoală, păscut, cresc, crescător, crescînd, descoperi, descusut, risca, iscăli, iscodi, iscusit, mosc, cunoscut, usca, crepuscul; indecşi, şcolar, şcoală, taşca, mişca, mişcă, cloşca; vrednic; doctorand, sanda, bandă, candelă, candid, suspendat, endem, apendice, citînd, scînda, scîndez, scîndează, indispus, indus, văzînd, îndată, îndelete, pîndi, îndoi, îndura, rond, sonă, rondel, condiţie, bondoacă, fondul, rotund, funda, unde, ascundea, fundul; draperie, drept, dreaptă, drob, drum, adresă, cadru, clorhidric, hidrofor, codri; ard, arde, gardă, cardiac, pardoseală, verde, pierdut, fiord, mordant, cordeluţă, sordid, cordon, surda; flacăra, flămînd, floare, afla, află, afli, sufla, suflă, sufle; golf, sulf, sulfuric; frază, frăgezime, frenetic, frige, frînge, fruct, cifre, şofran; harfă, parfum, ferfeniţă, perfora, vîrf, vîrful, morfologie; ig-nora; angaja, gîngaş, stingă, stingînd, lingură, plîng, gîngav, plîngă, plîngînd, îngust, ung, lungă, lungă, lunguiet; gras, grăbit, greşi, gresea, grimasă, grîne, gros, groasă, grumaz, agrar, neagră, agreabil, agricol, agronom, integra, negre, înnegri, negru, digresiune, ogradă, progresa, zugrăvi; sparg, argat, argăţi, merg, tergal, cergă, mergînd, organic,

curg, murgul; talpă, alpin, stîlp, vulpea; plan, plăcut, pleca, pleacă, plin, plîns, plod, ploaie, aplană, aplica, deplin, tiplă, disciplină, diplomă, supliment; muls, pulsație, mulsă; slab, slănină, slobod, slugă, rîsle; alt, alta, daltă, alte, saltea, saltimbanc, palton, saltul, sălta, săltă, deltei, cheltui, bolta, boltă, bolte, boltit, mult, sultan, cultă, adulte, ultima, insultînd, tumultos, vultur; titlu; halva, alveolar, vîlvă, rezolva, rezolvă, rezolve; vlagă; încălzește; răzleț, dezlega; amnar, doamnă, doamne, semn, semna, semnă, semne, temniță, lemnul, schimnic, somnifer, somnul, dumnezeu, dumneavoastră, pumnii; înmormînta; mrea-nă; armaș, armă, parmezan, sărman, termal, ferme, terminal, infirm, infirma, firmă, infirmii, cîrmui, enorm, formal, formă, forme, formînd, urmaș, urmă, turme, urmînd; samsar; smîntînă, turism, idiotisme, optimismul; simtă, simte; atmosferă, ritm, aritmetică, ritmic; înrudit; toarnă, carne, carnea, carnivor, infern, hiberna, perne, vernisaj, cer-nută, bîrnă, stîrni, orna, bornă, borne, ceasornic, turna, urnă, turnesol, urni; ansamblu, expansiv, densă, densitate, tensiune, pensulă, prins, întinsă, prinse, insulta, plîns, gînsac, plînsă, însetat, înseamnă, în-sorit, însoară, însura, consacrat, consecvent, consistent, uns, unse; snop, lesne, înlesni, josnic; neant, panta, pantă, constante, antinomie, pantof, dantură, strident, inventa, lentă, lente, dentiție, absolutul, argint, înainte, înaintează, intim, vînt, cînta, cîntă, cînte, trîntea, întins, întîlni, întocmi, întoarce, întunec, contabil, contează, conti-nent, contor, contur, nunta, nuntă, punte, untură; etnic; cheniză, împînzi, pînza, brînză, pînze, vînzînd, prînzul, bronz, bronzul, frunză, frunze, tunzînd; praznic, trăznește, trăzni, bezna; prag, prăda, prepeli-ță, prea, prim, program, proastă, căpra, capre, aprig, aprod, aproape, căprar, căprui, depresiune, deprimant, reproș, opreliște, opri, supra, cuprins; harpa, cîrpe, cîrpeală, serpentină; lipsă, lipsi, oropsi; spa-nac, spătar, special, spital, spîn, sport, spune, proaspăt, aspect, răs-pîndi, răspăr, răspuns, despăgubi, lespede, despotic, nespus, ispăși, ispiti, disponibil, dispus, gospodar, suspenda, suspin; arta, artă, parte, cartea, artist, cartuș, depărta, depărtă, cert, certa, certe, hîrtii, port, sorta, aortă, portic, ortoped, mortul, purta, purtă, curtezan, curtînd, burtos; trage, trăda, tesar, treacăt, trimite, trînti, tropot, trupă, patra, atrăgînd, patrimoniu, patru, pătrat, pătrătea, pătruns, către, bătrîn, metraj, cronometră, vetre, metric, metronom, vitreg, potrivit, șotron; nerv, inerva, vervă, percers, inervează, servil, survenit; vraf, vreme, vreo, vrînd, manevre, necropat).

Pe baza acestor grupări, date de distribuția grupurilor conso-nantice simetrice de două elemente, vom stabili tipurile de simetrie

specifice limbii române pentru situația : vocală + consoană + consoană + vocală.

Datele obținute au fost introduse și de astă dată în matrice. Matricele includ grupările fonematice $\frac{C_1 C_2}{C_2 C_1}$ ³², în care liniile și coloanele reprezintă pe V_1 , respectiv V_2 . Matricele sînt de tipul :



Utilizînd aceleași criterii de constituire a matricelor, ca și pentru grupurile consonantice, și considerînd grupurile vocalice $V_1 V_2$, respectiv $V_2 V_1$, în contextele C_1 (care precedă grupul) — C_2 (care urmează grupul), rezultă următoarele cazuri :

$$[1] \begin{cases} V_1 C_1 C_2 V_2 \\ V_1 C_2 C_1 V_2 \end{cases}$$

$$[2] \begin{cases} V_1 C_1 C_2 V_2 \\ V_2 C_1 C_2 V_1 \end{cases}$$

$$[3] \begin{cases} V_1 C_2 C_1 V_2 \\ V_2 C_2 C_1 V_1 \end{cases}$$

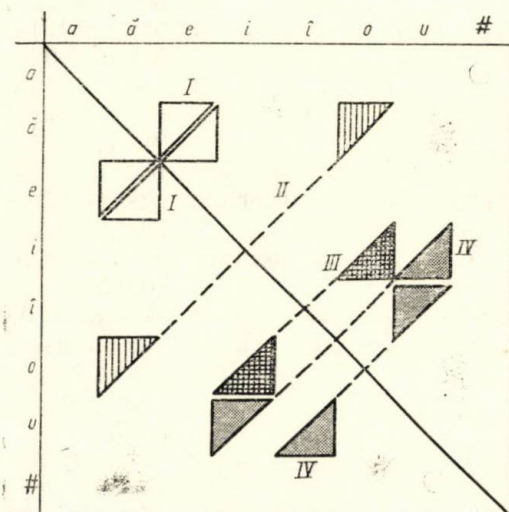
³² C_1 respectiv C_2 notează consoanele, iar V_1 respectiv V_2 notează vocalele.

$$[4] \begin{cases} V_1 C_1 C_2 V_2 \\ V_2 C_2 C_1 V_1 \end{cases}$$

$$[5] \begin{cases} V_1 C_2 C_1 V_2 \\ V_2 C_1 C_2 V_1 \end{cases} \quad (\text{care coincide cu } [4])$$

$$[6] \begin{cases} V_2 C_1 C_2 V_1 \\ V_2 C_2 C_1 V_1 \end{cases} \quad (\text{care coincide cu } [1])$$

În matrice, cele 4 cazuri sînt reprezentate astfel:



Pentru grupurile consonantice inventariate de noi, înregistrăm următoarele cazuri de simetrie în matricele:

- I *bl/lb* : [1] Cu prob. diferite : a bl a (0,025)³³/alba (0,075),
 ublu (0,025)/ulbu (0,05)
 [3] „ „ „ : ă lb a (0,05)/albă (0,125) ;
 [4] „ „ „ : # lb o (0,1)/olb# (0,07)
- II *br/rb* : [1] „ „ „ : abri (0,125)/arbi (0,025),
 obra (0,07)/orba (0,27) ;
 [2] „ „ „ : ibra (0,1)/abri (0,125) ;
 și „ „ identice : ubre/eburu (0,05) ;

³³ Procentele reprezintă frecvența grupurilor simetrice în contextele respective, în raport cu totalul de circa 6 000 grupuri simetrice înregistrate.

				[4] Cu prob. diferite :	##bro (0,45)/orb# (0,025) ; ibra (0,1)/arbi (0,025)
III	<i>bz/zb</i> :			Nici unul din cele 4 cazuri expuse mai sus.	
IV	<i>kl/lk</i> :			Idem	
V	<i>kn/nk</i> :			Idem	
VI	<i>kr/rk</i> :	[1]		Cu prob. diferite :	akră (0,025)/arkă (0,17), ek- ru (0,07)/erku (0,12), ukra (0,15)/urka (0,12) ;
		[2]	„	„	ukre (0,22)/ekru (0,07) ;
		[3]	„	„	urka (0,12)/arku (0,025) ; ărka (0,07)/arkă (0,17) ;
		[4]	„	„	ukre (0,22)/erku (0,12) ; uk- ra (0,15)/arku (0,025), ărka (0,07)/akră (0,025) ;
		și	„	„	identice :
VII	<i>ks/sk</i>	[1]	„	„	irka/akri (0,025) diferite : eksă (0,17) eskă (0,025), ekso (0,025)/esko (0,15) ;
		și	„	„	identice : aksa/aska (0,025), iksa/iska (0,025) ;
		[2]	„	„	iksa/aksi (0,025) ;
		[3]	„	„	##ska/ask# (0,025) ;
		și	„	„	diferite : ##sko/0,9/osk# (0,1), uska (0,17)/aksu (0,1), ășka (0,02) askă (0,75) ;
		[4]	„	„	##ska (0,25)/aks# (0,25), oksi (0,25)/ikso (0,025) ;
		și	„	„	identice : iska/aksi (0,025) ;
VIII	<i>kʃ/ʃk</i> :			Nici unul din cele 4 cazuri expuse mai sus.	
IX	<i>dn/nd</i> :	[1]		Cu prob. diferite :	edni (0,025)/endi (0,15) ;
		[3]	„	„	inda (0,1) andi (0,12), inde (0,5)/endi (0,15), enda (0,17) /ande (0,1) ;
		[4]	„	„	inde (0,5)/edni (0,025)
X	<i>dr/rđ</i> :	[1]	„	„	adre (0,12)/arde (0,2), odri (0,025)/ordī (0,1) ;
		[2]	„	„	odri (0,02)/idro (0,05) ;
		[3]	„	„	orda (0,05)/ardo (0,12) ;
		[4]	„	„	##dra (0,3)/ard# (0,025), ##dro (0,05) ord# (0,1), ur- da (0,12)/adru (0,025), idro (0,05)/ordi (0,17) ;

XI	<i>fl/lf</i> : [4]	„	„	identice :	#flo/olf # (0,05)
XII	<i>fr/rf</i> : [4]	„	„	diferite :	#fri (0,17)/irf # (0,025)
XIII	<i>gn/ng</i> :	Nici unul din cele 4 cazuri expuse mai sus.			
XIV	<i>gr/rg</i> : [1]	Cu prob. identice :			
					agra/arga (0,025), erga/egra (0,025) ;
	și	„	„	diferite :	agră (0,07)/argă (0,05), ogra (0,3)/orga (0,7) ;
	[2]	„	„	identice :	igre/egri (0,025)
	și	„	„	diferite :	egra (0,025)/agre (0,1) ;
					ogra (0,3)/agro (0,025),
	[4]	„	„	„	: #gra (0,4)/arg # (0,025),
					#gre (0,22)/erg # (0,025),
					#gro (0,17)/org # (0,05),
					orga (0,7)/agro (0,02), erga (0,025)/agre (0,1) ;
XV	<i>lp/pl</i> : [1]	„	„	„	: alpi (0,17)/apli (0,05) ;
	[4]	„	„	„	: #pli (0,22)/îlp # (0,025)
XVI	<i>ls/sl</i> [4]	„	„	„	: uls # (0,025)/#slu (0,17)
XVII	<i>lt/tl</i> [2]	„	„	„	: ulta (0,35)/altu (0,1), ulte (0,4)/eltu (0,025) olta (0,55)/alto (0,025), ălta (0,025)/altă (0,4) ;
	[4]	„	„	„	: ulți (0,3)/itlu (0,025) ;
XVIII	<i>lv/vl</i> :	Nici unul din cele 4 cazuri expuse mai sus.			
XIX	<i>lz/zl</i> : [1]	Cu prob. identice : ălze/ăzle (0,025) ;			
XX	<i>mn/nm</i> : [2]	„	„	diferite :	umne (0,15)/emnu (0,025), emna (0,12)/amne (0,025)
XXI	<i>mr/rm</i> : [3]	„	„	identice :	urmî/îrmu (0,025) ;
	și	„	„	diferite :	erma (0,075)/arme (0,17), arma (0,075)/armă (0,05) ;
XXII	<i>ms/sm</i> :	Nici unul din cele 4 cazuri expuse mai sus.			
XXIII	<i>mt/tm</i> [1]	Cu prob. identice : imte/itme (0,05)			
XXIV	<i>nr/rn</i> : [3]	„	„	diferite :	urni (0,075)/îrnu (0,025), erma (0,075)/arne (0,025) ;
XXV	<i>ns/sn</i> [1]	„	„	„	: ensi (0,12)/esni (0,075), onsi (0,3)/osni (0,025) ;
	[2]	„	„	„	: unse (0,22)/ensu (0,05), inse (0,22)/ensi (0,12) ;
	[4]	„	„	„	: inse (0,22)/esni (0,075)

XXVI	<i>nt/tn</i>	[1]	„	„	„	:	enti (0,1)/etni (0,025);
		[2]	„	„	„	:	unta (0,75)/antu (0,02), en- ta (0,75)/ante (0,3), unte/en- tu (0,15), onta (0,2)/anto (0,1), inta (0,075)/anti (0,3), inte (1,25)/enti (0,1);
		[4]	Cu	prob.	diferite	:	inte (1,25)/etni (0,025)
XXVII	<i>nz/zn</i>	[2]	„	„	identice	:	unzî/inzu (0,025)
XXVIII	<i>pr/rp</i>		„	„	diferite	:	apără (0,025)/arpă (0,075), epre (0,17)/erpe (0,03);
		[4]	„	„	identice	:	opre/erpo (0,025)
XXIX	<i>ps/sp</i>	[1]	„	„	„	:	ipsă/ispă (0,025), ipsi/ispî (0,025);
		[3]	„	„	diferite	:	uspi (0,025)/ispu (0,1), ospî (0,15)/ispo (0,25).
XXX	<i>rt/tr</i>	[1]	„	„	„	:	arta (0,1)/atra (0,125), artă (0,125)/atră (0,025), arti (0,7)/atri (0,3), artu (0,025) atru (0,5) ărtă (0,125)/ătra (0,025), ărtă (0,025) /atră(0,17), erta (0,075)/ /etra(0,025), erte (0,025) /etre(0,05), orti (0,125) /otri(0,75);
		[2]	„	„	„	:	arte (0,075)/erta (1,17), ar- tu (0,025)/urta (0,1), ărtu (0,025)/(0,31);
	și	„	„	„	identice	:	artă/ărtă (0,125);
	[3]	„	„	„	„	:	atră/ătra (0,025);
	și	„	„	„	diferite	:	ătre (0,15)/etră (0,025), itre (0,05)/etri (0,2);
	[4]	„	„	„	„	:	arte (1,1)/etra (0,025), urta (0,1)/atru (0,5), ert#(0,025) /#tre (19,75), ort#(0,2)/# tro (0,1)
XXXI	<i>rv/vr</i>	[1]	„	„	„	:	erve (0,075)/evre (0,025);
		[4]	„	„	identice	:	erv#/#vre (0,025) —

Pe baza acestor date, putem sintetiza frecvența diverselor tipuri de simetrie în raport cu contextele date. În tabloul următor se ilustrează această situație:

Grup de cons.	Tip [1]		Tip [2]		Tip [3]		Tip [4]	
	%		%		%		%	
	ident.	dif.	ident.	dif.	ident.	dif.	ident.	dif.
bl/lb		a-a				ă-a		#-o #-o
br/rb		u-u a-i o-a	u-e	i-a				i-a
bz/zb kl/lk kn/nk kr/rk		a-ă e-u u-a e-ă e-o	i-a	u-e		u-a ă-a		u-e a-a i-a
ks/sk	a-# i-a				#-a #-o u-a ă-a		i-a	#-a o-i
k{/k dn/nd		e-i				i-a i-e e-a o-a		i-e
dr/rd		a-e o-i		o-i				#-a #-o u-a i-o #-i
fl/lf fr/rf gn/ng gr/rg	a-a e-a	a-ă o-a	i-e	o-a e-a			#-o	#-a #-e #-o o-a e-a #-i u-#
lp/pl ls/sl lt/tl		a-i		u-a u-e o-a ă-a				
lv/vl lz/zl mn/nm	ă-e			u-e e-a	u-î	e-a ă-a		
mr/rm ms/sm nt/tm	i-e							

Grup de cons.	Tip [1]		Tip [2]		Tip [3]		Tip [4]	
	%		%		%		%	
	ident.	dif.	ident.	dif.	ident.	dif.	ident.	dif.
nr/rn						u-i i-a		
ns/sn		e-i o-i		u-e i-e				i-e
nt/tn		e-i		u-a o-a e-a i-a u-e i-e				i-e
nz/zn			u-i					
pr/rp		a-ă	o-e				o-e	
ps/sp	i-ă i-î					u-i o-i		
rt/tr	a-a a-ă a-i a-u ă-a ă-ă e-a e-e o-i			a-e a-u ă-u	a-ă	ă-e i-e		a-e u-a e-# o-#
rv/vr		e-e					e-#	o-#

1.4.2. Definim ca *grupuri consonantice asimetrice* grupurile consonantice care prezintă o singură orientare posibilă, prin permutarea elementelor lor nemaierezultând grupuri proprii limbii române. Dintr-acele 10 000 de grupuri consonantice cu 2 elemente inventariate, circa 3 000 sînt grupuri asimetrice, deci 33% din totalul grupurilor consonantice cu două elemente.

Grupurile consonantice asimetrice înregistrate sînt următoarele : *bă, kă, km, kt, kă, dm, dv, ft, gă, gl, hn, 3b, 3d, l, 3lă, lă, lm, ln, mb, mp, mă, nă, nf, nă, n3, nă, nv, pt, pă, ră, ră, rh, r3, rl, rs, ră, rt, ră, rz, sf, st, řk, řn, řt, zd, zg, zl, zm.*

Grupurile de mai sus diferă între ele atît prin frecvență, cît și prin posibilitatea de distribuire într-un număr mai mare sau mai mic de contexte diferite.

Tabelul de la p. 79 reflectă frecvențele grupurilor consonantice metrice.

Din acest tabel reiese că grupurile consonantice asimetrice cele mai frecvente sînt *st* (22,06), *řt* (8), *nă* (7,85); cel mai puțin frecvente sînt *kv, gă, gn, řk, zg, zm* (0,05).

Grup. cons.	%	Grup. cons.	%
bd	0,1	nğ	3,2
kč	0,25	nž	0,3
km	0,6	nt	7,85
kt	6,7	nv	1,3
kŧ	1,65	pt	4,4
kv	0,05	pŧ	0,5
dm	0,45	rč	1,5
dv	0,65	rğ	2
ft	0,6	rh	0,05
gd	0,05	rž	0,05
gl	1,1	rl	0,15
gn	0,05	rs	2,3
hn	2,6	rš	1,15
žb	0,05	rŧ	0,9
žd	0,25	rz	0,5
žl	0,6	sf	2,55
lč	0,1	st	22,06
lğ	0,1	šk	0,05
lm	0,55	šn	0,25
ln	0,75	št	8
mb	2,2	zd	0,1
mp	6,25	zg	0,05
mŧ	0,25	zl	0,1
nč	5,7	zm	0,05
nf	4,3	zv	0,75

Lista completă ³⁴ a contextelor vocalice în care se distribuie grupurile consonantice cu două elemente asimetrice este următoarea :

bd : ă-a

kč : u(e, i)

km : o(a, i)

kt : a(#, e, i, o u), e(- #, a, ă, e, i, o, u), i(-e, i, o, u), o-o, u(-o, u)

kŧ : a-u

kv : e-e

dm : a-i

³⁴ În această listă am introdus numai exemplele înregistrate de noi în textele parcurse. Am făcut abstracție de contexte cunoscute în limbă care, neapărând în textele excerptate, denotă o mai mică probabilitate de apariție. De exemplu, pentru *bŧ* : a-i (*abŧine*), *kv* : a-a (*acvatie*), *dm* : a-o (*admonesta*), *žb* : u-ă (*slujbă*), *šn* : a-i (*pașnic*) etc.

ŷt : #(-a, e, i,), a(# -e, i, ă,), ă-i, e(-a, ă, i), i(-e, i), i-i, î-i, o(-ă, e, i)

zd : a-e, u-u

zg : # -o

zl : ă-e, e-e

zm : # -u

zv : # -î, e(-ă, o), i-o.

Grupurile consonantice considerate se repartizează deci în următoarele numere de contexte :

Grup cons.	Repartiția (la nr. de contexte)	Grup cons.	Repartiția (la nr. de contexte)
bd	1	nğ	9
kè	2	n3	4
km	2	nț	17
kt	19	nv	7
kț	1	pt	23
kv	1	pț	3
dm	1	rè	6
dv	1	rh	1
ft	6	r3	1
gd	1	rl	1
gl	1	rs	11
gn	1	rș	3
hn	5	rț	12
3b	3	rz	4
3d	2	sf	13
3l	1	st	42
lè	2	șk	2
lğ	1	șn	2
lm	4	șt	16
ln	6	zd	2
mb	11	zg	1
mp	20	zl	2
mț	3	zm	1
nè	15	zv	4
nf	11		

Exemple : / răbda ; sukčes, sukčint ; tokmai, tokmi ; akt, akte, aktiv, aktor, paktul, selekt, selekta, korektă, aspekte, korektează, kolektiv, korektor, aspektul, oktombrie, induktor, fruktul ; akțiune ; konsekvent ; administra ; adverb ; rafturi, ofta, poftă, pofte, oftikos

mofturos; amigdale; glas, glesnă, glosar, glumă, aglomerat, regula, reglementar, oglindi; ignora; tehnik, tehnoeditor, tihnă, lihnit, mîhnit; slu3ba, slu3bă, slu3be; prime3die, mu3dei; mi3lok, mi3loace; ulcele, melci; mulge; kalm, kalma, kalmă, psalmi, ulm, kulmea; vremelnik, silnik, întilnesk, întilni, bolnav, îmbolnăvi; plimb, plimba, limbă, skimbe, imbold, strîmbă, strîmbe, îmbogăți, îmbunătăți, plombă, kombinat; ștampilă, kampion, krampona, temperatură, timp, impas, limpede, imposibil, impune, kimp, împarte, împărți, împiedika, kimpul, kompara, kompetent, komponent, kompus, skump, skumpă, skumpi; zimț, zimțat, simți; bocanči, lanče, lănci, enčiklopedie, činči, incendiu, inčita, brinči, inčetini, končedia, končis, atunci; a3unge; infarkt infekt, infinit, infuzie, înfățișa, infipt, infumurat, konfekție, konfiguratie, konformatie, konfuz; borančik, stingi, inženios, inžiner, plinče, konželat, unge, ungi, lungi; in3ekta, oglin3oară, in3unghia, inkon3ur; lanț, klanța, klanță, klanțe, lanțul, lănțișor, kadența, kadență, kadențe, torențial, știința, kredință, ființe, părinți, înțeles, renunța, punțile; anvelopă, inventa, învăța, învekit, învins, konvins, konvoka; apt, adapta, faptă, șapte, faptul, săptămîna, drept, aștepta, îndreptăți, trepte, șceptik, așteptînd, dreptul, eliptik, opt, inopta, inoptă, șoptește, șopti, rupt, lupta, suptă, lupte; rečeptie, nopțile, suptire; inkarči, inkarče, inčerka, čerčetez, kokostirči, urči; arčilă, marġini, merġe, merġea, ștergi, skurġi, kurġe, urġisi; orhidee; perlat, perlă, perle, burlan; bir3ă, ars, arse, mers, persan, ștersătură, șterse, persona3, persoană, mersul, întors, urs, ursită, kursul; arȳče, sfirȳște, sfirȳit; despărți, arțar, parțial, kărțulie, terțiar, komerțul, șorț, forță, forțe, porți, porție, kurți; inverzi, dirz, întirzi, urzeală; sfat, sfătui, sfetnik, sfinks, sfint, asfinți, desfače, desfăkut, resfira, desfunda, satisfakție, atmosferă, fosfor; sta, stă, stele, stiklă, stîlp, stop, stup, asta, pastă, kastel, astîmpăr, bastoane, ăsta, răsti, răsturna, rest, kresta, vestă, peste, pastilă, bestie, estompa, destul, turist, lista, listă, liste, turistik, istovit, artistul, rost, kosta, foste, rosti, rostogoli, rostul, gust, gusta, fustă, pustiu, gustînd, gustoase, gustul; mušk, škiop; vešnik, obișnui; ștafetă, șterge, știre, poștă, kaște, așteaptă, praștie, măști, inkleșta, pește, pești, liniște, liniștea, optimiști, kiștig, poște, oștean, proști, fokiști; brazde, buzdugan; zgomot, răzleț, dezlega; zmulge; zvikni; dezvălui, dezvolta, izvorî (*răbda; succes, succint; tocmăi,*

toemi; act, acte, activ, actor, pactul, select, selecta, corectă, aspecte, corectează, colectiv, corector, aspectul, octombrie, inductor, fructul; acțiune; consecvent; administra; adverb; rafturi, oftă, poftă, poște, ofticos, mofturos; amigdale; glas, gleznă, glosar, glumă, aglomerat, regula, reglementar, oglindă; ignora; tehnic, tehnoredactor, tihnă, lihnit, mîhnit; slujba, slujbă, slujbe; primejdie, mujdei; mijloc, mijloace; ulcele, melci; mulge; calm, calma, calmă, psalmii, ulm, culmea; vremelnic, silnic, întîlnesc, întîlni, bolnav, îmbolnăvi; plimb, plimba, limbă, schimbe, imbold, strîmbă, strîmbe, îmbogăți, îmbunătăți, plombă, combinat; ștampilă, campion, crampona, temperatură, timp, impas, limpede, imposibil, impune, cîmp, împarte, împărți, împiedica, cîmpul, compara, competent, component, compus, scump, scumpă, scumpi; zîmț, zîmțat simți; bocanci, lance, lănci, enciclopedie, cinci, incendiu, incita, brînci, încetini, concedia, concis; atunci; ajunge; infarct, infect, infinit, infuzie, înfățișa, înfipt, înfumurat, confecție, configurație, conformație, confuz; borangic, stingi, ingenios, inginer, plînge, congelat, ungi, unge, lungi; injecta, oglinjoară, înjunghia, înconjur; lanț, clanța, clanță, clanțe, lanțul, lăntîșor, cadența, cadență, cadențe, torențial, știința, credință, ființe, părinții, înțeleș, renunța, punțile; anvelopă, inventa, învăța, învechit, învins, convins, convoca; apt, adapta, faptă, șapte, fapta, săptămîna, drept, aștepta, îndreptăți, trepte, sceptic, așteptînd, dreptul, eliptic, opt, înnopta, înnoptă, șoptește, șopti, rupt, lupta, suptă, lupte; recepție, nopțile, subțire; încarce, încarci, încerca, cercetezi, cocostîrci, urci; argilă, margini, merge, mergea, ștergi, scurgi, curge, urgisi; orhidee; perlat, perlă, perle, burlan; birjă; ars, arse, mers, persan, ștersătură, șterse, personaj, persoană, mersul, întors, urs, ursită, cursul; arșice, sfîrșește, sfîrșit; despărți, arșar, parțial, cărțuție, terțiar, comerțul, șorț, forță, forțe, porții, porție, curți; înverzi, dîrz, întîrzi, urzeală; sfat, sfătui, sfetnic, sfinx, sfînt, asfinți, desface, desfăcut, resfîra, desfunda, satisfacție, atmosferă, fosfor; sta, stă, stele, sticlă, stîlp, stop, stup, asta, pastă, castel, astîmpăr, bastoane, asta, răști, răsturna, rest, creșta, vestă, peste, pastilă, bestie, estompa, destul, turist, lista, listă, liste, turistic, istorit, artistul, rost, costa, foste, roști, rostogoli, rostul, gust, gusta, fustă, pusti, gustînd, gustoase, gustul; mușc, șchiop; veșnic, obișnui; ștafetă, șterge, știre, poștă, caște, așteaptă, praștie, măști, înclășta, pește, pești, liniște, liniștea, optimiști, cîștig, poște, oștean,

proști, fochiști; brazde, buzdugan; zgomot; răzleț, dezlega; zmulge; zvîeni; dezvălui, dezvolta, izvorî).

Pe baza grupărilor date de distribuția grupurilor consonantice asimetrice de 2 elemente, vom stabili tipurile de simetrie specifice limbii române.

Pentru acest tip de grupuri introducem din nou notația [$\rightarrow$] (simbolul acesta fiind utilizat în general ca element-ventil, element de unică orientare posibilă), care indică sensul în care este posibilă citirea grupului.

Grupurile consonantice asimetrice sînt de tipul :

$$\begin{cases} V_1 C_1 \rightarrow C_2 V_2 \\ V_2 C_1 \rightarrow C_2 V_1 \end{cases}$$

Specific e faptul că în acest caz se pot face comparații doar pe baza contextelor vocalice, întrucît se pot permuta numai vocalele, grupul consonantic rămînînd invariabil.

Ca și în studiul grupurilor consonantice simetrice, pentru a sistematiza prezentarea materialului analizat, am introdus datele obținute în matrice, $C_1 C_2$ (care cuprind și procente de apariție) în care liniile și coloanele reprezintă pe V_1 , respectiv V_2 .

Pentru grupurile inventariate de noi, înregistrăm următoarele contexte simetrice ³⁵ :

Matrice <i>kt</i> :	<i>a</i> — <i>e</i> (<i>a kt e</i> = 0,3/ <i>e kta</i> = 0,85)
	<i>e</i> — <i>i</i> (<i>e kt i</i> = 1,05/ <i>i kt e</i> = 0,05)
<i>mb</i> :	<i>i</i> — <i>o</i> (<i>i mb o</i> = 0,05/ <i>o mb i</i> = 0,45)
	<i>mp</i> :
	<i>a</i> — <i>i</i> (<i>a mp i</i> = 0,35/ <i>i mp a</i> = 0,1)
	<i>a</i> — <i>e</i> (<i>a mp o</i> = 0,05/ <i>o mp a</i> = 0,45)
<i>nč</i> :	<i>e</i> — <i>i</i> (<i>e nč i</i> = 0,05/ <i>i nč e</i> = 0,1)
	<i>nț</i> :
	<i>a</i> — <i>e</i> (<i>a nț e</i> = 0,15/ <i>e nț a</i> = 0,40)
	<i>a</i> — <i>i</i> (<i>a nț i</i> = 0,15/ <i>i nț a</i> = 0,25)
	<i>a</i> — <i>u</i> (<i>a nț u</i> = 0,1/ <i>u nț a</i> = 0,3)
	<i>e</i> — <i>i</i> (<i>e nț i</i> = 0,50/ <i>i nț e</i> = 0,05)
<i>pt</i> :	<i>a</i> — <i>e</i> (<i>a pt e</i> = 0,05/ <i>e pt a</i> = 1,05)
	<i>e</i> — <i>u</i> (<i>e pt u</i> = 0,15/ <i>u pt e</i> = 0,1)
	<i>a</i> — <i>u</i> (<i>a pt u</i> = 0,15/ <i>u pt a</i> = 0,05)
	<i>rs</i> :
	<i>a</i> — <i>e</i> (<i>a rs e</i> = 0,3/ <i>e rs a</i> = 0,1)

³⁵ Observăm că în toate cazurile notate probabilitățile de apariție diferă între ele.

<i>sf</i> :	a — i (a sf i = 0,05/ i sf a = 0,10)
<i>st</i> :	# — e (#st e = 0,35/ e st# = 1,65)
	# — i (#st i = 1,15/ i st# = 0,80)
	# — o (#st o = 0,10/ o st# = 2,2)
	# — u (#st u = 0,80/ u st# = 0,05)
	a — ă (a st ă = 0,40/ ă st a = 0,15)
	a — e (a st e = 0,15/ e st a = 0,60)
	ă — i (ă st i = 0,15/ i st ă = 0,45)
	ă — u (ă st u = 0,05/ u st ă = 0,10)
	a — e (a st e = 0,15/ e st a = 0,60)
	e — i (e st i = 0,95/ i st e = 1,10)
	e — e (e st o = 0,40/ o st e = 0,25)
	i — o (i st o = 0,25/ o st i = 0,05)
	i — u (i st u = 0,25/ u st i = 0,20)
<i>ŝt</i> :	a — e (a ŝt e = 1,4/ e ŝt a = 0,05)

Un caz particular este cel în care grupul asimetric C_1C_2 , contextul vocalic V_1V_2 apare sub forma $V_1 \equiv V_2$.

Am înregistrat această situație pentru cazurile următoare :

Matrice :	<i>kt</i> :	e-e (0,20), i-i (0,15), u-u (0,1)
	<i>kv</i> :	e-e (0,05)
	<i>gl</i> :	e-e (0,05)
	<i>hn</i> :	i-i (0,12)
	<i>lm</i> :	a-a (0,2)
	<i>ln</i> :	i-i (0,05)
	<i>mp</i> :	e-e (0,45), o-o (0,2)
	<i>mŝ</i> :	i-i (0,15)
	<i>nč</i> :	i-i (0,50)
	<i>nf</i> :	i-i (0,2), o-o (0,15)
	<i>nğ</i> :	i-i (0,2)
	<i>nŝ</i> :	a-a (0,05), e-e (0,2), i-i (0,75)
	<i>nv</i> :	o-o (0,11)
	<i>pt</i> :	a-a (0,1), e-e (0,2), i-i (0,05)
	<i>rč</i> :	e-e (0,1)
	<i>rğ</i> :	e-e (0,45)
	<i>rs</i> :	e-e (0,2), u-u (0,40)
	<i>st</i> :	a-a (0,90), e-e (2,85), i-i (0,55), u-u (0,05)
		o-o (0,05)
	<i>ŝt</i> :	e-e (2,80), i-i (0,45)
	<i>zd</i> :	u-u (0,05).

Contextele care apar, evident, cel mai frecvent în această situație, sînt *e-e*, *i-i*, ceea ce este explicabil, dacă avem în vedere probabilitatea mare a vocalelor respective, în limba română.

1.4.3. În limba română există și o serie de segmente complexe mai largi, de trei consoane, de tipul $C_1C_2C_3$.

Din materialul excerptat de noi au rezultat următoarele grupuri cu structura aceasta :

ksč, ksp, kst, ktr, ltč, lft, ltm, ltr, mbl, mbr, mfl, mpl, mpr, mps, mpt, nkl, nkr, nkt, nkč, ndr, ndv, nfl, nfr, ngr, nml, nsf, nsm, nsp, nst, nšf, nšt, ntk, ntr, psr, psč, pst, rpr, rst, rtč, rtr, skl, skr, sfr, spl, spr, stf, stm, str, štr, zbr, zdr, gzl, zgn ³⁶.

Analizînd aceste grupuri din punctul de vedere al structurii lor consonantice, notăm consoanele din grupurile de trei elemente — de la stînga —, după locul (întîi, al doilea și al treilea) pe care îl ocupă un grup, și obținem următoarea situație :

- locul I : p, k, s, z, ʃ, m, n, l, r ;
- locul II : p, b, t, d, k, g, f, s, ʃ, m ;
- locul III : p, t, d, k, ʈ, č, f, v, s, m, l, r.

Locurile I, II, III se definesc, pentru consoane, astfel : spunem că o consoană ocupă locul I atunci cînd într-o secvență de trei poate fi urmată de altele două, iar imediat după ea urmează măcar una dintre consoanele de pe locul II ; spunem că o consoană ocupă locul II atunci cînd într-o secvență de trei poate fi precedată măcar de una dintre consoanele de pe locul I și urmată măcar de una de pe locul III ; spunem că o consoană ocupă locul III atunci cînd într-o secvență de trei ea este precedată de două consoane și urmează măcar unei consoane de pe locul II ³⁷.

Rezultă că :

— Primul loc poate fi ocupat de : oclusivele *p* și *k* ; constrictivele *s*, *z*, *ʃ* ; nazalele *m*, *n* ; laterala *l* și vibranta *r*.

Nu poate fi ocupat de constrictivele *f*, *v*, *ʒ*, *h*.

³⁶ Ca și în cazul grupurilor de două consoane, în limba română mai sînt posibile și alte grupuri de trei, patru și cinci consoane, care n-au apărut însă în materialul parcurs de noi. Probabilitatea lor de apariție fiind, deci, sub 1 la totalul materialului, ele au fost neglijate în inventarul nostru. Asemenea grupuri sînt, de exemplu : *-ršt* (*crenvršt*), *-stm* (*istm*), *-ždr* (*mojdrean*), *-lkl* (*folelor*), *-nks* (*sfinx*) etc.

³⁷ La fel ca pentru grupurile vocalice ternare, considerăm necesară o descriere a structurii grupurilor consonantice ternare.

— Locul al doilea este ocupat de : oclusivele *p, b, t, d, k, g*; constrictivele *f, v, s, ʃ*; nazalele *m, n*.

Nu poate fi ocupat de : semioclusive, laterale, vibrante.

— Locul al treilea este ocupat de : oclusivele *p, t, k*; semioclusiva *č*; constrictivele *f, v, s*; nazala *m*; laterala *l*; vibranta *r*.

— Sînt excluse de pe orice loc din grupurile de trei consoane : oclusivele *k', g'*; semioclusiva *ǵ*; constrictivele *ʒ* și *h*.

În tabelele ³⁸ care urmează înregistrăm frecvența acestor grupuri și distribuția lor.

Grup. cons.	%	Grup. cons.	%
ksč	0,69	nsm	0,05
ksp	0,5	nsp	0,6
kst	0,5	nst	0,8
ktr	0,10	nšf	0,5
ltč	0,2	nšt	0,8
ltf	0,8	ntk	0,06
ltm	0,06	ntr	27,9
ltr	0,1	psk	0,1
mbł	1,2	psč	0,06
mbr	1,64	pst	0,06
mfl	0,8	rpr	0,4
mpl	9	rst	0,25
mpr	9,1	rtf	0,06
mps	0,06	rtr	0,15
mpt	0,06	skł	0,10
nkl	1,25	skr	3,1
nkr	1,4	sfr	0,10
nkt	0,25	spl	0,28
nkč	0,66	spr	5,5
ndr	2,4	stf	0,8
ndv	0,25	stm	0,1
nfl	0,67	str	17,4
nfr	0,38	štr	1,29
ngl	0,06	zbr	0,2
ngr	1,4	zdr	0,2
nml	0,1	zgl	0,1
nsf	0,2	zgr	0,02

³⁸ Frecvența de apariție este frecvența grupurilor consonantice în raport cu totalul grupurilor înregistrate de noi. Repartiția reflectă distribuția fiecărui grup în cele 52 de contexte vocale distincte care au rezultat din cercetarea noastră.

Din aceste tabele rezultă că cele mai frecvente grupuri consonantice cu trei elemente sînt: *ntr* (28,4), *str* (17,4), *mpl*, *mpr* (9,1).

Lista completă a contextelor vocalice în care se distribuie grupurile consonantice cu trei elemente este următoarea:

ksč: e(-e, i); *ksp*: e-e; *kst*: e(-#, a, e, i); *ktr*: a-i; *ltč*: a-e; *ltf*: a-e; *ltm*: a-i; *ltr*: a-u, u-a; *mbł*: a-a, i-i, u(-a, ă, i); *mbr*: e(-e, i), i-i, i(-a, ă), i-ă, u(-a, ă, e, o); *mfl*: u(-a, ă, u); *mpl*: a-o, e(-i, u), i(-ă, e, u), i(-a, ă, e, i), o(-a, e, i), u(-e, i, u); *mpr*: i(-e, i, o, u), i(-ă, e, o, u), o(-i, o); *mps*: e-i; *mpt*: i-o; *nkl*: i(-e, i), o-u; *nkr*: i(-e, o, u), o-e; *nkt*: i(#, i, u), u-#; *nkț*: u-i; *ndr*: a-u, i(-ă, e, i, u), *ndv*: i-a; *nfl*: i(-e, u), o-i, i-o; *nfr*: i-i, o-a; *ngl*: a-i; *ngr*: i-a, i(-ă, e, i, o); *nml*: i-ă; *nsf*: a(-i, o); *nsm*: a-i; *nsp*: a-o, i-i, i-ă; *nst*: i(-a, e, i), i-e, o(-a, ă, i); *nsf*: i-a; *nst*: o(-i); *ntk*: i-ă; *ntr*: a(-a, e, o), e(-e, u), i(-a, ă, e, i, u), i(-a, ă, e, i, i, o, u), o(-a, o), u(-i, u); *psk*: o-u; *psč*: o-e; *pst*: o-a; *rpr*: e-e, u-i; *rtf*: e-e; *rtr*: o-e; *rst*: i(-a, ă); *skl*: #-i; *skr*: #(-i, u), ă-u, e-i, i-e; *sfr*: ă-i, e-i, ă-ă, i-ă, e-e; *spr*: #(-e, i, i), a(-e, i, u), ă-e, e(-e, i), i(-ă, e); *stf*: a-e; *stm*: a-a; *str*: #(-a, ă, e, i, i, o, u), a(-a, ă, e), e(-a, ă, e, i, u), i(-a, e, u), o(-o, u), u(-a, ă, u); *str*: #-e, a-i, o-i; *zbr*: e(-a, ă); *zdr*: #-a; *zgl*: #-o; *zgr*: #-i.

Grupurile consonantice se distribuie, deci, în următoarele numere de contexte:

Grup. cons.	Repartizare (la nr. de contexte)	Grup. cons.	Repartizarea (la nr. de contexte)
ksč	2	mps	1
ksp	1	mpt	1
kst	4	nkl	4
ktr	1	nkr	4
ltč	1	nkt	3
ltf	1	nkț	1
ltm	1	ndr	6
ltr	2	ndv	1
mbł	5	nfl	4
mbr	10	nfr	2
mfl	3	ngl	1
mpl	16	ngr	5
mpr	10	nml	1

Grup. cons.	Repartizare (la nr. de contexte)	Grup. cons.	Repartizare (la nr. de contexte)
nsf	2	rtr	1
nsm	1	skl	1
nsp	3	skr	5
nst	7	sfr	2
nŝi	1	spl	3
nŝt	2	spr	12
ntk	1	stf	1
ntr	3	stm	1
psk	1	str	24
psĉ	1	ŝtr	3
pst	1	zbr	2
rpr	2	zdr	1
rst	2	zgl	1
rtf	1	zgr	1

Exemple : ³⁹/eksĉeptie, eksĉitat; eksperienĉă; tekst, ekstaz, tekste, pretekstind; aktriĉă; altĉeva, altfel, altminteri, altruist, ultraĉu; ansamblu, imblinzi, umbla, umblă, umblind, dezmembrează, septembrie, limbrik, imbrakă, imbrăkat, umbra, umbră, umbre, umbros; umfla, umflă, umflu; amployare, kontemplind, kontemplu, simplă, simple, simplu, timplar, timplă, timple, implini, komplaĉe, komplet, komplika, umple, umpli, umplut; impregna, imprima, impropriu, imprudent, imprăŝtia, impreĉmui, improspăta, imprumuta, komprima, kompromis; metempsihoză; simptome; inkleŝta, inkлина, konkluzie; inkredinĉa, inkropi, inkrusta, konkret; instinkt, instinktiv, instinktu, punkt; funkĉiona; tandru, mindră, mindre, mindrie, mindru, mindra; kindva; infleksioni, influks, konflikt, inflori; infrinĉe, konfrate; panglikă; ingrat, ingrăŝa, ingreuia, ingriĉora, ingropa; inmlădia; transfigura, transforma; transmite; transport, inspira, inspăiminta; instabil, ĉinste, ĉinsti, instelat, konstata, konstă, constitui; inŝăka; konŝtiinĉă, konŝtient; fiindkă; antraĉit, antrenor, antropofag, ĉentre, ĉentru, kontrakt, kontrol, lăuntriĉ, lăuntru; opskur, opsĉen; opstakol; interpreta, surpriză; virsta, virstă; ĉertfe; portret; sklipi; skrisoare, skrum, răskruĉe, deskris, diskrepanĉă; răsfrinĉe, desfriu; răsplăti, displăkut, despleti; spre, spriĉin, sprinĉene, aspre, asprime, aspru, nouăsprezeĉe, despre, des-

³⁹ Ca ŝi în cazul grupurilor consonantice am introdus în această listă numai exemple înregistrate de noi în textele parcurse, făĉind deci abstrakĉie de alte contexte cunoscute în limbă. De exemplu : *kst* : e-i (*extins*), *mpl* : a-u (*amplu*) etc.

prins, isprăvi, dispreț, paisprezece; astfel; astmatik; stradă, străbun, strepezi, streașină, striga, strînge, stropi, strugure, astral, albăstri, astre, destramă, destrămat, zestrea, pestriț, distructiv, distra, distră, distrează, distruge, ostrov, nostru, lustragiu, lustră, lustrui; ștengar, aștri, noștri; dezbracă, dezbrăca; zdrăvăn; zglobiu; zgripturoaică/ (excepție, excitat; experiență; text, extaz, texte, pretextînd; acțiță; altceva; altfel; altminteri; altruist, ultragiu; ansamblu, îmblînzi, umbla, umblă, umblind; dezmembrează, septembrie, limbric, îmbracă, îmbrăcat, umbra, umbră, umbre, umbros; umfla, umflă, umflu; amployare, contemplînd, contemplu, simplă, simple, simplu, tîmplar, tîmplă, tîmple, împlini, complace, complet, complica, umple, umpli, umplut; impregna, imprima, impropriu, imprudent, imprăstia, împrejmu, improspăta, împrumuta, comprima, compromis; metempsihoză; simptome; încesta, înclina, concluzie; încredința, încropi, încrusta, concret; instinct, instinctiv, instinctul, punct; funcționa; tandru, mîndră, mîndre, mîndrie, mîndru, mîndra; cîndva; inflexiuni, influx, conflict, înflori; înfrînge, confrate; panglică; ingrat, îngrășa, îngreuiă, îngrijora, îngropa; înmlădia; transfigura, transforma; transmite; transport, înspira, înspăimînta; instabil, cinste, cinsti, înstelat, constata, constă, constitui; înșfăca; conștiință, conștient; fiindcă; antracit, antrenor, antropofag, centre, centru, contract, control, lăuntric, lăuntru; obscour, obscen; obstacol; interpreta, surpriză; vîrsta, vîrstă; jertfe; portret; scilipi; scrisoare, scrum, răscruce, descris, discrepantă; răsfrînge, desfrîu; răsplăti, displăcut, despleti; spre, sprijin, sprîncene, aspre, asprime, aspru, nouăsprezece, despre, desprins, isprăvi, dispreț, paisprezece; astfel; astmatic; stradă, străbun, strepezi, streașină, striga, strînge, stropi, strugure, astral, albăstri, astre, destramă, destrămat, zestrea, pestriț, distructiv, distra, distră, distrează, distruge, ostrov, nostru, lustragiu, lustră, lustrui; ștengar, aștri, noștri; dezbracă, dezbrăca; zdrăvăn; zglobiu; zgripturoaică).

Din examinarea listei contextelor vocalice în care pot apărea aceste grupuri, rezultă că:

- în poziție inițială apar: skl, skr, spr, str, ștr, zdr, zgl;
- în poziție finală apar: kst, nkt;
- toate grupurile pot apărea în poziție medială.

Matricele care sistematizează materialul analizat sînt de tipul $C_1C_2C_3$, în care liniile și coloanele reprezintă pe V_1 și V_1 — contextele vocalice ale grupurilor consonantice respective.

Pentru situația: $V_1C_1C_2C_3V_2$ putem stabili două tipuri de simetrii, pe baza contextelor vocalice:

a) simetria care rezultă din posibilitatea permutării elementelor vocale, grupul consonantic rămânând invariabil;

b) simetria în care contextul vocalic V_1V_2 apare sub forma (particulară) $V_1 \equiv V_2$.

Cazurile inventariate de noi sînt următoarele :

- a) Matrice *ltr* : a-u (altru = 0,62/ ultra = 0,062)
 Matrice *mpl* : e-u (emplu = 0,54/umple = 0,126), i-u
 (implu = 0,062/umpli = 0,56)
 Matrice *mpr* : i-o (impro = 0,062/ompri = 0,062)
 Matrice *skr* : e-i (eskri = 0,062/iskre = 0,25)
 Matrice *spr* : e-i (espri = 0,124/ispre = 0,56)
- b) Matrice *ksč* : e-e (0,68)
ksp : e-e (0,062)
kst : e-e (0,124)
mbl : î-î (0,126)
mbr : i-i (0,062)
mfl : u-u (0,124)
mpl : u-u (0,68)
mpr : i-i (0,124), o-o (0,27)
mkt : i-i (0,062)
nfr : î-î (0,5)
ntr : a-a (0,062), e-e (0,12), i-i (0,31), î-î (0,12),
 o-o (0,12), u-u (0,43)
nsp : i-i (0,43)
nst : i-i (0,126)
rpr : e-e (0,062)
rtf : e-e (0,062)
spl : ă-ă (0,124), e-e (0,124)
spr : e-e (0,68)
stm : a-a (0,124)
str : a-a (0,54), e-e (0,37), u-u (0,126).

Din materialul parcurs de noi, au rezultat următoarele grupuri de patru consoane, de tip CCCC și de cinci consoane, de tip CCCCC⁴⁰ *kskl*, *kspl*, *kspr*, *kstr*, *mskr*, *nčspr*, *nskr*, *nstr*, *nspr*, *pspr*, *rstm*, *stskr*.

⁴⁰ Din grupurile consonantice inventariate de noi, circa 96,39% sînt grupuri consonantice cu trei elemente; 2,96 cu patru elemente; 0,40 cu cinci elemente.

Frecvența și distribuția acestora sînt următoarele :

Grup. cons.	%	Grup. cons.	%
kskl	0,06	nċspr	0,06
kspl	1,25	nspr	0,31
kspr	0,48	nstr	0,40
kstr	0,25	ptspr	0,18
mskr	0,06	rstn	0,12
nskr	0,06	stskr	0,16

Aceste grupuri, destul de rare în limba noastră, nu prezintă diferențe sensibile ca frecvență și în modul lor de repartizare în diverse contexte.

Grupurile cu mai mult decît două consoane se reduc de fapt, adeseori, în rostire.

De exemplu : [altminter^d] > [alminter^d], [astmatik^h] > [asmatik^h], [eksklama] > [esklama] etc.

De aceea, numărul de elemente considerat de noi se referă de fapt la o pronunțare „ideală”. Am recurs la acest criteriu, pentru că nu puteam „prevedea” toate reducțiunile posibile.

Dacă interpretăm (ca și în cazul grupurilor vocalice cu trei elemente) consoana medială drept „axă de simetrie”, secvențele cu structuri simetrice sînt : *rpr*, *rtr*; cele în care consoanele extreme au o poziție „răsturnată” față de axa de simetrie sînt : *ksp* / *psk*.

Contextele în care se distribuie aceste grupuri sînt : *kskl* : e-a; *kspl* : e(i-o); *kspr* : e(e, i); *kstr* : e(a, e); *mskr* : u-i; *nċspr* : i-e; *nskr* : i-i; *nspr* : i-e, u-e; *nstr* : i-u, o-u; *ptspr* : o-e; *rstn* : î-i; *stskr* : o-i.

Fiecare grup consonantic se repartizează, deci, în următoarele numere de contexte :

Grup cons.	Repartiție (la nr. de contexte)	Grup cons.	Repartiție (la nr. de contexte)
kskl	1	nskr	1
kspl	1	nspr	2
kspr	2	nstr	2
kstr	2	ptspr	1
mskr	1	rstn	1
nċspr	1	stskr	1

Exemple : / eksklama, eksplika, eksplozie, ekspresie, eksprima ; ekstrage, ekstremă, ċirkumskrie, ċinċsprezeċe ; inskriptċie ; dinspre, unsprezeċe ; instruktor, konstruktor ; optsprezeċe ; vċirstnik ; postskriptum / (exclama, explica, explozie, expresie, exprima ; extrage, extremă ; ċircumserie ; ċincisprezece ; inscriptċie ; dinspre, unsprezece ; instructor, constructor ; optsprezece ; vċirstnic ; postscriptum).

Contextele simetrice (în cazul de faċă formate numai din vocale identice) rezultă din matricele $C_1C_2C_3C_4(C_5)$:

Matrice *kstr* : e-e (0,06)
kspr : e-e (0,43)
nskr : i-i (0,12).

1.5. Finala cuvintelor

Am cercetat *finala* cuvċntului, ċnċelegċnd prin aceasta numai ultimul fonem dinainte de pauză, indiferent de faptul că fonemul respectiv constituie singur o silabă sau este doar un element constitutiv al unei silabe ori al unui morfem. Am avut ċn vedere numai cuvintele cu mai mult decċt o silabă și am făcut investigaċii pe texte scrise din limba română actuală.

ċntr-o descriere generală observăm că aproape toate fonemele limbii române pot apărea ċn poziċie finală, după cum urmează : / halva, ramă, krab, kozonak, șubred, perete, taraf, belșug, almanah, dori, hotări, baraș, oral, volum, elan, domino, sirop, fecor, kolos, paloș, urċt, drăguċ, roșu, elev, nekaz / (*halva, ramă, crab, cozonac, șubred, perete, taraf, belșug, almanah, dori, hotări, baraj, oral, volum, elan, domino, sirop, fecior, colos, paloș, urċt, drăguċ, roșu, elev, necaz*).

Studiul statistic pune ċn lumină următoarele frecvenċe pentru aceste foneme :

a — 14,2, *ă* — 14, *b* — 0,01, *k* — 1,39, *d* — 1,45, *e* — 28,18, *g* — 0,10, *i* — 9,97, *ș* — 0,16, *l* — 7,46, *m* — 1,60, *n* — 1,77, *o* — 0,33, *p* — 0,07, *r* — 5,35, *s* — 1,23, *ș* — 0,20, *t* — 8, *ț* — 1,03, *u* — 2,50, *v* — 0,21, *z* — 0,19.

Se observă că unele foneme : *f*, *h*, *î* nu apar de loc ċn tabelul de frecvenċe, ocurenċa lor fiind deci 0 (zero).

Fonemele care apar cel mai frecvent ċn poziċie finală sċnt cele vocalice. 6 vocale totalizează 69,02 din finalele cuvintelor luate ċn consideraċie, iar 16 consoane reprezintă 30,07 ċn aceeași situaċie.

Așezċnd ċn ordinea descrescċndă a frecvenċelor fonemele românești dispuse comparativ, după procentele de ocurenċă ċn toate pozi-

țiile și după frecvența la finală, obținem un tabel care reflectă următoarele ⁴¹:

Ocurență în toate pozițiile	Ocurență la finală	Ocurență în toate pozițiile	Ocurență la finală
e	e	i	ɪ
a	a	l	
î	ă	p	o
n	i	v	v
r	t	ʃ	ʃ
u	l	f	z
ă	r	ț	ʒ
t	u	b	g
o	m	g	p
s	d	z	
m	n	h	b
k	k	ʒ	
d	s		f, h, i

În primul rînd observăm diferența de pondere a vocalelor în situațiile comparate. Față de 69,02% din totalul finalelor, reprezentat de 6 vocale (a 7-a, *î*, neapărînd în tabelul nostru) în cazul care reprezintă totalul celorlalte poziții în limbă, frecvența totală a vocalelor este 46% (incluzînd și pe *î*, cu 3,4% ea este de 49,4%). Este evidentă preferința limbii pentru utilizarea vocalelor la finală (ne referim la ansamblul vocalelor limbii române).

Extinzînd comparația la situația fonemelor românești în poziție inițială (*vezi capitoulul respectiv*), observăm că *ă*, care la inițială ocupă rangul cel mai mare (poziția ultimă în tablou), cu 0,1%, la finală are rangul III și frecvența 14%. Într-o situație opusă se află *î*, care este a doua vocală cu frecvență mare (6,5%) la inițială (rangul V), iar la finală nu a fost înregistrat de loc, în statistica noastră.

Vocalele care au procente cele mai mari și rangurile cele mai mici — *e*, *a*, *i* — se aflau la ranguri apropiate și în cazul tota-

⁴¹ Am considerat utilă o comparație între diverse poziții, întrucît ea poate oferi distincții cantitative specifice, atît între limbi, cît și în interiorul aceleiași limbi. Astfel dintre datele pe care le avem la dispoziție, în limba engleză se observă că vocalele alcătuiesc numai 1/3 din finalele cuvintelor (totalul frecvențelor lor însumînd 39,27% ca poziție generală în limbă). Vocale frecvente în poziție finală, sînt cele frecvente și la inițială, și în limbă în general. Consoanele cele mai frecvente în limbă în alte poziții nu sînt însă frecvente și la finală Cf. P.B. Deneș, *On the Statistics of Spoken English*, p. 58—59.

lului pozițiilor din limbă, iar ca procente prezintă cifre asemănătoare (i final = 9,97 / față de 8,1 a final = 14,2 / față de 8,8); o diferență de procente sensibilă apare numai în cazul lui e (e final = 28,18 / față de 10,9).

În ceea ce privește consoanele, cele mai frecvente la finală sînt t , l , r ; în cazul totalului pozițiilor din limbă, r ocupă al doilea rang dintre consoane (și rangul V în tabel); în poziție inițială acestea figurează la ranguri mai mari, în cazul respectiv primele locuri din tabel aparținînd lui k , s , p . Compararea procentelor, care în cazul lui t variază între 5,42% și 8%, în cazul lui l , între 3,1% și 7,46%, în cazul lui r , între 7,5% și 5,35%, nu ne permite, cel puțin deocamdată, stabilirea unei reguli.

O ultimă observație privește configurația tabelelor de frecvență pentru situațiile considerate mai sus. Pentru cazul care reprezintă ocurența în toate pozițiile, în primele 3 ranguri se află 3 vocale (e , a , i); tot 3 vocale (e , a , $ă$), ocupă și primele 3 ranguri în poziție finală. Se poate menționa că, în cazul inițialelor, în primele 3 ranguri se află consoanele.

Frecvențele deosebite ale fonemelor, într-un caz sau într-altul, conduc la întrebarea: care este cauza acestor diferențe? Dacă pentru inițiale s-au dat pînă acum explicații mai ales diacronice⁴², în cazul finalelor vom urmări în primul rînd o descriere sincronică.

Este de observat faptul că, în limba română, mai importante pentru înțelegerea relațiilor dintre cuvinte în frază sînt elementele care se adaugă la sfîrșitul lor, urmînd rădăcinii. Morfemele⁴³ finale pot cuprinde una sau două silabe și se încheie cu un fonem final al cărui rol îl vom explica în cele ce urmează.

Rămîne să urmărim dacă există vreun raport între informația gramaticală — raportul dintre expresie și conținut (funcția-semn) — pe care o poartă aceste foneme și frecvența lor.

Fonemele finale românești pot marca, respectiv, următoarele raporturi⁴⁴:

⁴² Cf. Al. Graur, *Încercare asupra fondului principal lexical al limbii române*, p. 55—57.

⁴³ Pentru definiția morfemului, vezi *Gramatica limbii române*, I, p. 42.

⁴⁴ Trebuie să avem în vedere, de altfel, importanța cunoașterii valorilor morfologice și, eventual, sintactice ale fonemelor, într-o limbă dată. Cf. și: «Nous pensons que toute statistique de phonèmes doit tenir compte de valeurs significatives de ceux-ci». Intervention de B. Pottier [în] *Actes du sixième Congrès international des linguistes*, Paris, juillet, 1948, p. 87.

- a : — vb. ind. imperf. 3 /putea/ (putea)
inf. /verifika/ (verifica)
— subst. fem. art. Nom. sg. /frunza/ (frunza)
- ă : — vb. ind. prez. 3, 6 /sună/ (sună)
imperf. 3 /repetă/ (repetă)
perf. simplu 6 /intrară/ (intrară)
m.m. ca perf. 6 /clădiseră/ (clădiseră)
subj. prez. 3,6 /străbată/ (străbată)
participiu /uskată/ (uskată)
— subst. fem. neart. Nom. sg. /platoșă/ (platoșă)
— adj. fem. neart. Nom. sg. /bună/ (bună)
- k : — vb. ind. prez. 3 /pornesk/ (pornesc)
subj. prez. 1 /găsesc/ (găsesc)
— adj. masc. Nom. sg. /românesk/ (românesc)
- d : — vb. gerunziu/kîntînd/ (cîntînd)
- e : — vb. ind. prez. 3 /taçe/ (tace)
perf. simplu 3 /zise/ (zise)
m.m. ca perf. 3 /venise/ (venise)
subj. prez. 3,6 /aștepte/ (aștepte)
imper. /spune/ (spune)
inf. /strînge/ (strînge)
— subst. fem. neutru neart. Nom. pl. /brațe, pauze/
(brațe, pauze)
art. /motivele/ (motivele)
— adj. fem. neart. Nom. /negre/ (negre)
art. /frumoasele/ (frumoasele)
- i : — vb. ind. prez. 2 /auzi/ (auzi)
perf. simplu 3 /citi/ (citi)
imperativ /rămii/ (rămîi)
infinitiv /muri/ (muri)
— subst. masc., fem., neutru, neart. Nom. pl. /metri/
(metri)
masc. art. /tineri/ (tinerii)
fem. art. Gen. sg. /apei/ (apei)
— adj. masc., fem., N. pl. /roși/ (roșii)
G. sg. /bunei/ (bunei)
- l : — subst. masc. art. Nom. sg. /cerul/ (cerul)
— adj. /frumosul/ (frumosul)

- *m* : — vb. ind. prez. 4 /servim/ (*servim*)
 imperf. 1,4 /aflam/ (*aflam*)
 m.m. ca perf. 1,4 /stabilisem/ (*stabilisem*)
 subj. /vorvim/ (*vorvim*)
- *o* : — subst. fem. art. Voc. sg. /fato/ (*fato*)
- *r* : — subst. masc., fem., neutre, art. Gen. pl. /cetăților,
 koleților, veakurilor/ (*cetăților, colegilor,*
 veacurilor)
 — adj. masc., fem. art. Gen. pl. /albelor, noilor/ (*albelor,*
 noilor)
- *t* : — vb. participiu /citit/ (*citit*)
- *u* : — vb. ind. imperf. 6 /tăiau/ (*tăiau*)
 perf. simplu 6 /treku/ (*treku*)

Finalele cuvintelor dau indicații gramaticale pentru diferite aspecte flexionare, așa cum rezultă din exemplele de mai sus. În cercetarea statistică pe care am efectuat-o se „atestează” majoritatea acestor situații. În afară de cazurile când un fonem aflat în poziție finală marchează un raport gramatical, există o serie de situații când fonemele respective apar fără rol gramatical (formele din flexiunea unor cuvinte alcătuite din rădăcină sau temă la care nu se adaugă nimic, când avem a face cu sufixul gramatical zero sau desinența zero ⁴⁵).

Este evident însă că *e, i, a, ă* — vocalele care apar cel mai des — „cumulează” o serie de funcții gramaticale prin care se poate explica frecvența lor ridicată la finala cuvintelor. Pe de altă parte, frecvența mare a lui *t, l, r, u*, foneme care indică fiecare câte una sau cel mult două forme gramaticale, nu poate fi explicată decît fără referire la informația gramaticală pe care o conțin. Trebuie să putem înțelege, de asemenea, de ce *î*, care marchează infinitivul conjugării a IV-a, nu apare niciodată în tabelul nostru de frecvențe.

Vom observa că frecvența mare a unor foneme la finala cuvintelor se „motivează” printr-unul din elementele următoare :

- a) gramatical — mai ales pentru fonemele din primele ranguri *e, i, a, ă*, care indică, fiecare, o serie de raporturi;
- b) frecvență mare la finală — predictabilă întrucîtva atunci cînd știm că frecvența fonemului respectiv este mare și pentru limbă

⁴⁵ Cf. *Gramatica limbii române*, I, p. 44.

în general — în toate pozițiile (cazurile lui *t* — de exemplu, care e numai semn al indicativ perfectului compus și al modului participii; al lui *l* care este numai articol, al lui *r* care face parte numai din desinența substantivelor la genitiv plural */-lor/*). Și într-un asemenea caz se pot face însă distincții: un sondaj ne arată că, în timp ce frecvența mare a lui-*r* se bazează mai ales pe frecvența lui mare în limbă, pentru frecvența mare a lui-*l* avem ca explicație și marea frecvență a categoriei gramaticale pe care o desemnează acest fonem, articolul final. Astfel, din totalul de foneme-*r* numai 1/3 fac parte din desinența *-lor*, deci, numai acestea se justifică gramatical; pentru *-l* însă, din totalul de ocurențe, 8/10 au rolul de articol. Vom putea spune, așadar, că *-l* nu marchează decât o categorie gramaticală, dar aceea apare frecvent (cel puțin la o primă cercetare);

— *c*) frecvență mare la finală — explicabilă prin ocurența fonemului respectiv în poziția dată — într-un cuvânt (sau o serie de cuvinte) cu frecvență mare în limbă. Legătura strânsă dintre frecvența fonemelor la finală și diversele raporturi gramaticale pe care le redau este mai evidentă atunci când comparăm finalele unităților lexicale din dicționar (cuvintele-titlu) și finalele cuvintelor excerptate din texte (cuvintele-text).

Am efectuat astfel un sondaj în DLRM (aproximativ câte o pagină, ultima, a fiecărei litere din dicționar) și am obținut o serie de date pe care le expunem, încadrându-le într-un tabel comparativ (vezi p. 99).

Acesta cuprinde: finalele cuvintelor în texte și finalele cuvintelor în dicționar, ordonate după criteriul frecvenței și, respectiv, al rangului.

Din acest tabel rezultă următoarele:

În ceea ce privește rangul, fonemele finale au uneori în ambele situații (texte și dicționar) ranguri apropiate. Pentru a pune în evidență această situație, am împărțit tabelul în 3 secțiuni, prima și ultima incluzând câte 7 ranguri, a 2-a, 8 ranguri.

În secțiunea I se includ 6 din 7 foneme finale comparate (*e*, *a*, *ă*, *i*, *t*, *r*) și se deosebesc între ele numai 1 fonem (*l* care, de altfel, figurează între primele 2 ranguri ale secțiunii a II-a). În secțiunea a II-a se includ 6 din cele 9 foneme comparate (*u*, *m*, *d*, *n*, *s*, *ț*) și diferă ca rang numai 2 foneme (*k*, *o*). În secțiunea a III-a, din 7 foneme comparate, 2 au ranguri asemănătoare (*v*, *b*), iar 5 diferă (*ș*, *z*, *3*, *g*, *p*). Nu se includ în nici una dintre secțiuni următoarele foneme: 1 fonem (*h*) — care este același în ambele situații —

Rang	Finala în text	%	Finala în dicționar	%
1	e	28,18	ă	21,6
2	a	14,2	e	12,9
3	ă	14	ț	11,9
[I] 4	l	9,97	l	9,1
5	t	8	r	9,09
6	l	7,46	a	8,2
7	r	5,35	k	6,3
8	u	2,50	n	3,8
9	m	1,60	l	3,7
10	d	1,45	u	3,5
11	n	1,77	s	2,5
12	k	1,39	m	1,9
[II] 13	s	1,23	ș	1,29
14	ț	1,03	d	0,57
15	g	0,33	g, ȝ, p, ȝ, v	0,4
16	v	0,21		
17	ș	0,20		
18	z	0,19		
19	ȝ	0,16		
20	g	0,10	f, v	0,2
[III] 21	p	0,7	b	0,18
22	b	0,01	î	0,09
	f, h, î	0	h	0

--- Frecvențe asemănătoare ; ○ Frecvența cea mai mare
 □ Frecvența cea mai mică

și 2 care diferă (*f*, *i*, în texte). În total, deci, din cele 25 de foneme considerate, 14 au ranguri apropiate (aici poate fi inclus și al 16-lea fonem, *h*, notat peste tot cu 0%), iar 8 diferite. Asemănarea este cea mai pregnantă în secțiunea I a tabelului (deci unde se află foneme cu frecvențele cele mai mari).

În ceea ce privește frecvența, o primă deosebire apare când comparăm relațiile dintre vocale și consoane în cele două situații: în timp ce în texte ponderea cea mai mare este a vocalelor (vocale = 69,02%, consoane = 30,07%), în dicționar, majoritatea finalelor de cuvânt aparține consoanelor (consoane = 58,91, vocale = 45,59). Acest raport diferit duce la deosebiri în distribuția fonemelor, de la un caz la altul.

Frecvențe asemănătoare nu au decît 3 foneme, din fiecare secțiune cite unul: *i* (9,97/9,09), *m* (1,6/1,9), *o* (0,3/0,2). Pentru celelalte, variațiile sînt sensibile. Astfel, valori mai mari au în texte: *e*, *a*, *l*, *d*, *t*, *g*. În dicționar valorile mai mari sînt ale fonemelor: *ă*, *t*, *r*, *u*, *n*, *k*, *s*, *v*, *ș*, *z*, *3*, *p*, *b*.

Pentru unele foneme din această a doua categorie, care pot marca, după cum am arătat, și unele raporturi gramaticale, explicația frecvenței lor mari în dicționar este următoarea: *ă*, *t*, *r*, *u*, *k* — reprezintă și desinențe ale unor forme de bază ale cuvintelor, adică, tocmai ale cuvintelor-titlu pe care le înregistrează dicționarele (subst. *casă*, citit, *culegător*, *codru*, adj. *frumoasă*, *strălucitor*, românesc etc.). Diferențele mai pot proveni, pentru *t*, *r*, *k* și din frecvențele mai mari pe care am văzut că le au consoanele finale în dicționar. În asemenea cazuri se manifestă, de altfel, și legătura dintre frecvența în limbă, în general, a unui fonem și frecvența lui specială, într-un caz anumit, finala de exemplu.

În cazul în care frecvențele mai mari ale fonemelor sînt în dicționar, cît și în cazul cînd sînt în text, marcante sînt *deosebirile* generale între cele două situații distincte: text/dicționar. (Iată cîteva exemple mai grăitoare: raportul text/dicționar: $l = 2/1$, $d = 2\frac{1}{2}/1$, $n = 1/2$, $k = 1/5$, $t = 2\frac{1}{2}$, $ș = 1/6$, $3 = 1/4$, $b = 1/10$ etc.).

La „inventarul” de situații înregistrate, mai trebuie să adăugăm cîteva considerații. Astfel, cercetarea finalei cuvintelor impune, după observațiile referitoare la ultimele foneme care compun morfeme, remarci asupra alcătuirii acestora. Urmează să vedem, deci, ce raporturi se stabilesc în interiorul morfemelor și ce secvențe de foneme pot ele cuprinde.

Cunoscînd finala absolută, rămîne să stabilim în ce măsură putem „prevedea” fonemele care o precedă. De exemplu, pentru finala *-d* putem şti că secvenţa precedentă, caracteristica modului gerunziu, este: *cons.* sau *voc.* $+i +n$ $[+d]$; pentru *-t*, caracteristica modului participiu este *voc.* *a, i, î, u* $[+t]$ etc.

Mai multe posibilităţi de alegere apar în cazul fonemelor prin care se desemnează o serie de raporturi gramaticale (*e, i, a, ă*) şi, de asemenea, un număr mare de cazuri pentru finalele care sînt motivate gramatical (în cazul respectiv, singura indicaţie este aceea oferită de posibilitatea distribuţiei fonemelor în limbă).

O dată stabilită structura morfemelor, ar trebui să cunoaştem frecvenţa lor, şi, în acelaşi timp, frecvenţa diverselor forme şi categorii gramaticale, ceea ce ar depăşi cadrul cercetării noastre. Noi ne-am propus numai semnalarea unor aspecte ale studiului cantitativ în relaţiile pe care limba română le marchează prin *fonemele ultime* care apar la finala cuvintelor.

2. LUNGIMEA SECVENȚELOR FONEMATICE

2.1.0. Prin considerațiile pe care le-am făcut asupra organizării fonemelor în segmente complexe vocalice sau consonantice, am anticipat unele elemente de mărime (lungime) a structurilor lingvistice.

Am stabilit că grupurile vocalice pot cuprinde pînă la patru vocale, iar grupurile consonantice — pînă la cinci consoane. În general, cu cît crește numărul elementelor de același fel (fie vocale, fie consoane), scade frecvența tipului de secvență pe care acestea o compun).

De astă dată, ne vom opri la descrierea unor unități de alt tip — silabele. Cercetarea structurii silabice cuprinde descrierea posibilităților de trecere de la vocale (sau grupuri vocalice) la consoane (sau grupuri consonantice), ori de la consoane (sau grupuri consonantice) la vocale (sau grupuri vocalice).

2.1.1. Silabele

Vom începe prin a delimita silabele fonologice, pentru a putea inventaria tipurile de silabe mai frecvente din limba română.

Silaba — definirea și delimitarea ei — stau în centrul atenției cercetătorilor, ca elemente indispensabile pentru analiza unei serii de fapte lingvistice¹.

Prin *silabă* se înțelege, în lucrarea de față, cea mai mică tranșă fonică susceptibilă de a fi caracterizată printr-un accent și numai unul².

¹ Cf. A. Rosetti, *Sur la théorie de la syllabe*, p. 11. Autorul expune în această lucrare și diverse puncte de vedere din care poate fi definită silaba.

² L. Hjelmslev, *The Syllable as a Structural Unit*, în *Proceedings of the Third International Congress of Phonetic Sciences*, p. 266—272.

Ca și în cazul altor unități fonematice, structura silabei diferă de la o limbă la alta. În limba română, silaba trebuie să conțină în mod obligatoriu o vocală.

Într-o descriere fonologică a limbii române, Em. Vasiliu³ analizează silaba, descriind secvențele fonematice construite prin alăturări succesive de segmente vocalice și segmente consonantice. Autorul indică de asemenea modalitatea de secționare a grupurilor intervocalice, avînd în vedere existența mai multor posibilități în anumite cazuri. Dificultatea, în ce privește diviziunea silabică, este împărțirea în silabe a grupurilor de consoane intervocalice; adeseori ele pot fi fragmentate în mai multe feluri, ceea ce face ca locul tăieturii silabice să nu poată fi determinat totdeauna univoc. După metoda propusă de L. Hjelmslev⁴ este util să se studieze inițiala și finala de cuvînt, care sînt considerate identice cu inițiala și finala silabei. Acceptînd că silabele din interiorul cuvîntului au același tratament cu silabele de la extremitățile lui, putem trece la împărțirea grupurilor mediale.

Am clasificat segmentele intervocalice inventariate în *Fonologia limbii române* avînd drept criteriu diferența între grupurile consonantice care se pot secționa într-un singur fel și cele care admit mai multe diviziuni. Pentru alcătuirea unui *inventar* complet al silabelor românești trebuie să ne stabilim de fiecare dată la cite o *singură* formulă de segmentare. De aceea, examinînd grupurile intervocalice care teoretic admiteau mai multe soluții (fiind analizabile în diverse componente) am adăugat considerațiile probabilistice propuse de O'Connor și Trimm⁵. Am comparat frecvența cu care apăreau segmentele componente⁶ în fiecare dintre variantele posibile și *am optat* pentru cazul reprezentat printr-o frecvență mai mare.

Aceste elemente ne oferă un *mod practic* de analiză a cuvintelor în silabe.

³ Emanuel Vasiliu, *Fonologia limbii române*.

⁴ L. Hjelmslev, *op. cit.*

⁵ J.D. O'Connor și J.L.M. Trimm, *Vowel, Consonant, and Syllable — a Phonological Definition*, p. 103—123: „The preference for one syllabic division as opposed to another may be explained in terms of the frequency of occurrence of different types of syllable finals and initials”, p. 121. Considerații statistice pentru despărțirea în silabe recomandă și P. Menzerath și W. Meyer-Eppler: „Wo die Silbentrennung zweifelhaft ist, wird man versuchen, für die verschiedenen Möglichkeiten eine getrennte statistische Untersuchung urchzuführen, um festzustellen, ob vielleicht einer der Fälle sich durch eine besonders einfache Gesetzmässigkeit auszeichnet”. *Sprachtypologische Untersuchungen*, p. 58.

⁶ Am avut în vedere circa 10 000 de grupuri consonantice (inițiale și finale).

b) Secvențe care pot fi segmentate *fie* într-un fel, *fie* într-altul :
 $VCCV = V + CCV$ sau
 $VC + CV$

Acestea sînt segmentele complexe în a căror structură se identifică segmente de tipul F + I și II, sau și II și F + I și FF. Lista lor completă este următoarea :

bl	br			3b	3d	3g	3m	3n
kl	kn	kr	kv	pn	pr	ps		
dr				sk	sf	sl	sp	st
fl	fr			șc	șk			
gl	gr			tr				
hl	hr			vr				
				zb	zm	zn.		

Avînd la dispoziție două posibilități teoretice de secționare a acestor secvențe, comparăm frecvența elementelor care intră în structura lor : urmează să stabilim dacă tăietura silabică va sta între cele două consoane sau înaintea grupului format din două consoane ¹¹.

Cercetarea faptelor pune în lumină situații distincte :

a) O serie de grupuri consonantice C_1C_2 apar mai rar (sau mult mai rar) decît consoanele C_2 , cu care le comparăm.

Segmentele la care ne referim sînt următoarele :

C_2	% de apariție	C_1C_2	% de apariție
b—	4,5	3b—	0 ¹²
		zb—	0
k—	8,45	sk—	5,82
		șk—	0
č—	2,26	șc—	0
d—	5,7	3d—	0
g—	2,1	3g—	0
m—	5,6	3m—	0

¹¹ Și aici și mai departe, comparațiile noastre vor rămîne exclusiv în domeniul structurii consonantice. Nu vom face referiri la vocale, pe de o parte pentru că discutăm segmentele consonantice, iar pe de altă parte, pentru că vocalele ar intra în discuție numai din perspectiva finalei de cuvînt. Or, într-o cercetare statistică asupra finalei cuvintelor (vezi *capitolul respectiv*) am stabilit situația specială a fonemelor românești în această poziție, ceea ce face necomparabile situațiile : finală de cuvînt / finală de silabă.

¹² Procentele 0 notează absența grupurilor respective în materialul excerptat de noi (10.000 de grupuri cons.).

C_2	% de apariție	$C_1 C_2$	% de apariție
n—	4	zm—	0,26
		kn—	0
		zn—	0
		pn—	0
		zn—	0
p—	9,8	sp—	6,21
s—	10,1	ps—	5,42
t—	3,9	st—	6,84
v—	3,2	kv—	0

Frecvențele indică posibilitatea de apariție a uneia dintre cele 11 consoane — care reprezintă pe C_1 în segmentele complexe, ca segment inițial — este în medie 5,41; probabilitatea de apariție a unuia dintre cele 17 grupuri consonantice $C_1 C_2$ este 1,44 — de asemenea ca segment inițial.

Vom decide în asemenea cazuri, în mod evident, ca locul tăieturii silabice să fie între cele două consoane, secvența urmînd să aibă, astfel, structura VC_1-C_2V .

b) Segmentele consonantice complexe $C_1 C_2$ în care C_2 sînt consoanele (lichidele) *l*, *r* apar, în total, cu o frecvență mai mare decît segmentele consonantice simple C_2 , anume *l*, *r*, cu care le comparăm. Aceste segmente sînt:

C_2	% de apariție	$C_1 C_2$	% de apariție
l—	3	bl—	0,82
		kl—	4,07
		fl—	2,03
		hl—	0
		pl—	6,84
		sl—	1,74
r—	2,7	br—	2,87
		kr—	5,92
		dr—	4,83

C ₂	% de apariție	C ₁ C ₂	% de apariție
		fr —	5,82
		gr —	7,30
		hr —	0,06
		pr —	21,05
		lr —	15,88
		vr —	5,26

(Situția este mai marcată pentru grupurile în care al doilea element este lichida *r*.)

Se observă că probabilitatea de a apărea una dintre cele două lichide, ca segment inițial (C₂), este 2,85; probabilitatea unuia dintre cele 16 grupuri consonantice C₁C₂, care au ca element C₂ o lichidă, este în medie 5,33, ca segment inițial.

Având în vedere acest grad de sudură al segmentelor complexe la care ne referim, considerăm că tăietura silabică le precedă, structura lor fiind deci: V — C₁C₂V.

3) Secvențele VCCCV — formate cu segmente complexe CCC — admit, de asemenea, mai multe posibilități de plasare a tăieturii silabice.

a) În cazul când segmentele —CCC— sînt identice cu un segment FFI, ca și în cazul când sînt identice cu un FFI și cu un FFF, *singura* împărțire posibilă este: VCCCV = VCC + CV.

Este cazul grupurilor consonantice următoare:

ksh
ldm lpt ltk ltč ltf ltm
mpş mpt
nkm nks nkş nčz nğz ndd ndv nsh nss ntl
ptn ptz
rkş rdm rdn rfş rpk rpş rtk rtf rth rtš rtm rtn rtp rts rtv
stf stm stn stp stv.

b) În cazul când segmentele —CCC— sînt identice cu un segment FII, *singura* împărțire posibilă este: VCCCV = VC + CCV.

Este cazul segmentelor consonantice următoare:

bgr
dgr
msp

nml nsd nzd nvr
 ppr pŝt
 svr
 tgr tŝt
 zvr.

c) În cazul cînd în structura secvenței VCCCV apare un —CCC— identic cu un FFI și cu un FII, ca și în cazul cînd identitatea este față de un segment FFI și un segment FII și cu un segment FFF, sînt posibile două diviziuni: VCCCV = VCC + CV sau VC + CCV

Lista acestor grupuri este următoarea:

ksp kst ktr
 lkl ltr
 mbl mbr mfl mpl mpr
 nkt nkŧ ndr ngl ngr ngv nŝf ntv
 pŝk pŝc pst
 rkt rgr rpl rpr rsp rst.

Și în cazurile de mai sus tipurile de segmentare depind de clasa din care fac parte elementele componente ale segmentelor complexe $C_1C_2C_3$. Vom compara din nou posibilitățile de plasare ale tăieturilor silabice: între C_1 de o parte și C_2C_3 de altă parte; sau între C_1C_2 de o parte și C_3 de partea cealaltă.

α) Pentru segmentele $C_1C_2C_3$ în care C_3 este o lichidă:

Segmentul analizat $C_1C_2C_3$		C_2C_3	% de apariție	C_3	% de apariție
ktr,	ltr	tr—	15,88	r—	2,7
mpl,	rpl	pl—	6,84	l—	3
mpr,	rpr	pr—	21,05	r—	2,7
ndr		dr—	4,83	r—	2,7
ntr		tr—	15,88	r—	2,7
mbi		bl—	0,82	l—	3
mbr		br—	2,86	r—	2,7
mfl		fl—	2,03	l—	3
ngl		gl—	2,03	l—	3
lkl		kl—	4,07	l—	3
ngr,	rgr	gr—	7,30	r—	2,7

În toate aceste grupuri probabilitatea lui C_3 ca inițială este 2,85, iar a segmentului C_2C_3 este 6,69, ceea ce ne face să optăm pentru o împărțire: $VC_1 - C_2C_3V$.

β) Pentru celelalte segmente $C_1C_2C_3$:

Segmentul analizat $C_1C_2C_3$	C_2C_3	% de apariție	C_3	% de apariție
ksp, rsp	sp—	6,21	p—	9,8
kst, pst, rst	st—	6,84	t—	3,9
nkt, rkt	kt—	0	t—	3,9
nkț	kț—	0	ț—	0,7
ngv	gv—	0	v—	3,2
nșf	șf—	0	f—	4
psk	sk—	5,82	k—	8,45
psč	sč—	0,03	č—	2,26

Cu privire la aceste secvențe, vom observa următoarele: *nkt*, *rkt*, *nkț*, *ngv*, *nșf*, *psč* — este indicat să se segmenteze în: $VC_1C_2 - C_3V$, întrucît segmentele complexe C_2C_3 care intră în componența lor au frecvențe 0 sau, în orice caz, sub unitate. (Chiar în DLRM, care inventariază unitățile lexicale ale limbii române, asemenea segmente apar în poziție inițială numai uneori și numai în cuvinte rare, precum: *clitorie*, *șfichi*.)

Pentru celelalte segmente, *ksp*, *rsp*, *kst*, *pst*, *rst*, *psk*, frecvențele de apariție indică, de asemenea, segmentarea $VC_1C_2 - C_3V$, întrucît probabilitatea unui segment complex C_2C_3 ca inițială este, în acest caz, în medie 6,62, iar cea a unui segment simplu C_3 , ca inițială, este în medie 7,35. Tăietura silabică va fi, deci, între C_2 și C_3 .

d) În cazul cînd segmentele —CCC— din secvența VCCCV se identifică cu un FII și cu un III, segmentul intervocalic poate fi divizat în două feluri:

$$VCCCV = VC + CCV$$

sau

$$V + CCCV$$

Aceste grupuri consonantice sînt:

spl spr

ștr

zgr

Comparația între segmentele inițiale posibile $C_1C_2C_3$ și C_2C_3 arată următoarele :

$C_1C_2C_3$	% de apariție	C_2C_3	% de apariție
spl—	0,03	pl—	6,84
spr—	3,73	pr—	21,05
ŝtr—	0,15	tr—	15,88
zgr—	0,03	gr—	7,30

Probabilitatea, în medie, ca un grup de 3 consoane $C_1C_2C_3$ să apară la inițială, este de 0,98, în timp ce probabilitatea ca un grup de 2 consoane C_2C_3 să apară în aceeași poziție este, în medie, 12,76. Evident, segmentarea pentru care vom opta este $VC_1-C_2C_3V$.

e) În cazul cînd segmentul intervocalic —CCC— se identifică și cu un FFI și cu un FII și cu un III, diviziunea silabică se poate efectua în trei feluri :

VCCCV = VCC + CV

sau

VC + CCV

sau

V + CCCV

Grupurile consonantice din această categorie sînt :

skl skr str.

Compararea frecvențelor indică următoarele :

C_3	% de apariție	C_2C_3	% de apariție	$C_1C_2C_3$	% de apariție
l—	3	kl—	4,07	skl—	0,06
r—	2,7	kr—	5,92	skr—	2,13
		tr—	15,88	str—	5,75

Rezultă că un segment C_3 inițial poate apărea în medie 2,85, unul C_2C_3 — 10,95, iar unul $C_1C_2C_3$ — 3,94. Desigur, vom opta din nou pentru segmentarea $VC_1-C_2C_3V$.

4) Secvențele construite cu segmente CCCC admit următoarele diviziuni :

a) Cînd sînt identice cu segmentele FFII, *numai* VCCCCV = VCC + CCV. Aceste sînt grupurile : *rddr, rggr*.

b) Cînd sînt identice cu FIII, *numai* VCCCCV = VC + CCCV. Această divizare se aplică grupului : *mskr*.

c) Cînd sînt identice și cu FIII și cu FFII, se pot diviza : VCCCCV = VC + CCCV

sau

VCC + CCV

Grupurile care fac parte din această categorie sînt :

kskr kspr

nstr

pstr.

Cercetarea lor probabilistică indică următoarele :

Grup de analizat $C_1C_2C_3C_4$	$C_2C_3C_4$	% de apariție	C_3C_4	% de apariție
<i>kskr</i>	<i>skr—</i>	2,13	<i>kr—</i>	5,92
<i>kspr, nspr</i>	<i>spr—</i>	4,73	<i>pr—</i>	21,05
<i>nstr</i>	<i>str—</i>	5,75	<i>tr—</i>	15,88

Probabilitatea medie de apariție a unui segment $C_2C_3C_4$ ca inițială este 4,20, pe cînd a unui segment C_3C_4 ca inițială este 14,28. Acest lucru este de așteptat, de altfel, dacă avem în vedere și unul din cazurile discutate mai sus, anume modurile în care putem fragmenta secvențele mediale *skr, spr, str*. Și în situația respectivă, grupurile acestea se divizau tot în C — CC¹³.

5) Segmentele intervocalice de forma CCCCC, identice numai cu segmente FFIII, cu care se construiesc secvențe VCCCCCV, nu se pot segmenta *decît* astfel : VCCCCCV = VCC + CCCV. Este cazul segmentului : *ptspr*.

Din cele expuse mai sus, se observă că delimitarea silabelor, pe baza fragmentării grupurilor consonantice, variază de la un caz la altul, astfel :

¹³ Opțiunea noastră diferă de data aceasta de indicațiile date în *Îndreptar orto-grafic, ortoepic și de punctuație*, p. 72, pe care nu le găsim justificate fonologic.

a) Pentru secvențele VCV și VCCCCCV — în cazurile extreme deci, cu cel mai mic și, respectiv, cel mai mare număr de segmente intervocalice — este posibilă câte o singură diviziune: $VCV = V + CV$. Deci: /masă = ma + să/ (*masă* = *ma* + *să*) $VCCCCCV = VCC + CCCV$ /cînăsprezece = cînă + sprezece/ (*cincisprezece* = *cinci* + *sprezece*)

Tăietura silabică în aceste grupuri este redundantă.

b) Pentru secvențele VCCV și VCCCCV, cazurile cu un număr par de segmente intervocalice, simpla informație asupra extensiunii segmentului intervocalic nu este suficientă. De la un grup intervocalic la altul trebuie să cunoaștem toate fonemele componente putînd stabili, în acest fel, tot câte o singură diviziune: $VCCV = VC + CV$ /hapsîn = hap + sîn/ (*hapsîn* = *hap* + *sîn*) $VCCCCV = VCC + CCV$ /ekskrescînță = eks + krescînță/ (*exkrescînță* = *ex* + *crescînță*).

Tăietura silabică se află în aceste cazuri exact la mijlocul segmentului consonantic complex. O situație aparte prezintă numai grupurile de consoane C_1C_2 în care C_2 este o lichidă și care se împart în: $V + CCV$ /akru = a + kru/ (*acru* = *a* + *cru*).

c) În cazul secvențelor VCCCCV avem de asemenea a face cu o situație specială. De la un grup consonantic ternar la altul — trebuind să cunoaștem din nou toate fonemele particulare — vom fixa limita silabică, rezultatul fiind:

$VCCCCV = \text{fie } VC + CCV$ /punktual = punk + tual/
(*punktual* = *punc* + *tual*)

$\text{fie } VCC + CV$ /umple = um + ple/ (*umple* = *um* + *ple*)

Deși atît în cazul b, cît și în cazul c rămîn posibile și în urma analizei noastre, câte două soluții diferite, deosebirea față de rezultatele cercetărilor anterioare constă în faptul că, pentru fiecare segment intervocalic în parte, putem determina cu precizie locul tăieturii silabice. De exemplu, știind că, în general, segmentele de tipul —CCC— se pot segmenta fie C + CC, fie CC + C, avînd informația că segmentul particular este *mbl*, divizarea va fi *m + bl*, iar dacă segmentul particular este *ksp*, divizarea va fi *k + sp*. Așadar, cele două posibilități privesc întreaga categorie discutată; fiecare segment în parte acceptă, în lumina datelor probabiliste, câte o singură modalitate de împărțire.

După ce am efectuat această analiză, putem stabili care sînt tipurile de silabe mai frecvente din inventarul limbii române, unitățile formate din vocale + consoane, în cadrul anumitor limite:



Metoda cantitativă, se dovedește utilizabilă cu succes, demonstrând o dată în plus că, alături de mijloacele instrumentale moderne, abordarea structurală a problemei, iar în cazul nostru cu mijloace statistice, poate să aducă noi clarificări¹⁴ într-una dintre cele mai disputate probleme — silaba.

2.1.2. Pe baza criteriilor de divizare a cuvintelor în silabe, stabilite mai sus, putem identifica o serie de structuri silabice care intră în componența cuvintelor românești¹⁵. Într-un material cuprinzând circa 10 000 de silabe (rezultând din diviziunea a circa 3770 de cuvinte, excerptate din texte diverse¹⁶, am înregistrat următoarele tipuri de silabe :

SILABE

<i>Inițiale</i>	<i>Mediale</i>	<i>Finale</i>
V /aič/, apukase/ (<i>aici</i> , <i>apucase</i>)	V /audă/ (<i>audă</i>)	V /bău/ (<i>bău</i>)
VV /iubite, uitat/ (<i>iubite</i> , <i>uitat</i>)	VV /băiatul/ (<i>băiatul</i>)	VV /aia, kuie/ (<i>aia</i> , <i>cuie</i>)
VVC /oaste/ (<i>oaste</i>)	---	VVC /graiul/ (<i>graiul</i>)
VC /armă/ (<i>armă</i>)	VC /înainte/ (<i>înainte</i>)	VC /kîte/ (<i>cîte</i>)
---	---	VCC /fiind/ (<i>fiind</i>)
CV /kîte/ (<i>cîte</i>)	CV /unikă/ (<i>unică</i>)	CV /haină/ (<i>haină</i>)
CVV /foaie/ (<i>foaie</i>)	CVV /domoala/ (<i>domoale</i>)	CVV /apusului/ (<i>apusului</i>)
---	---	CVVV /aveau/ (<i>aveau</i>)
CVVC /poartă/ (<i>poartă</i>)	CVVC /întoarse/ (<i>întoarse</i>)	CVVC /puteam/ (<i>puteam</i>)
---	---	CVVCV /însăimîntă- toare/ (<i>însăimîntă- toare</i>)

¹⁴ Cf. pentru această idee și B. Malmberg, *Remarks on a Recent Contribution to the Problem of the Syllable*, p. 8.

¹⁵ Ne referim la cuvintele de două și mai multe silabe. Cuvintele monosilabice au fost studiate în lucrarea Sandei Golopenția-Eretescu, *La structure phonologique des monosyllabes roumains*, la care ne raportăm atunci cînd comparăm silabele din cuvintele polisilabice cu cele monosilabice.

¹⁶ Ca și în alte cercetări ale noastre, materialul pe care îl avem în vedere sînt textele și nu cuvintele-titlu din dicționare.

<i>Inițiale</i>	<i>Mediale</i>	<i>Finale</i>
CVC /salva/ (<i>salva</i>)	CVC /arunkă/ (<i>aruncă</i>)	CVC /violet/ (<i>violet</i>)
CVCC/somptuos/ (<i>somptuos</i>)	---	CVCC /invers/ (<i>invers</i>)
CCV /klipa/ (<i>clipa</i>)	CCV /vitregă/ (<i>vitregă</i>)	CCV /afla/ (<i>afla</i>)
CCVV /spaimă/ (<i>spaimă</i>)	CCVV /distrează/ (<i>distrează</i>)	CCVV /păstrau/ (<i>păstrau</i>)
		CCVVC /viklean/ (<i>viclean</i>)
CCVC /frunte/ (<i>frunte</i>)	CCVC /kuprinsul/ (<i>cuprinsul</i>)	CCVC /suflet/ (<i>suflet</i>)
CCVVC /proaspăt/ (<i>proaspăt</i>)	---	---
CCCV /strada/ (<i>strada</i>)	---	---
CCCVC /strînse/ (<i>strînse</i>) ¹⁷		

Se observă că, potrivit unei descrieri generale, aceleași tipuri silabice pot apărea în poziție inițială, medială, finală, cu numai câteva excepții.

Fragmentării cuvintelor în silabe îi urmează însă o cercetare asupra posibilității concatenării silabelor, care va indica în ce măsură se repartizează silabele respective în pozițiile menționate. Din cele 10 000 de silabe incluse în cercetarea noastră am desprins 20 de tipuri distincte.

Structura acestora este următoarea :

Ca *dimensiune*, ele pot cuprinde între 1 și 5 foneme, deosebindu-se de monosilabe care pot cuprinde până la 7 foneme.

Cele mai scurte silabe sînt de tipul V, iar cele mai lungi de tipurile : CCVVC, CCVCC, CCCVC.

¹⁷ Din listele noastre lipsesc unele tipuri silabice posibile în limba română, dar pe care nu le-am întîlnit în materialul parcurs de noi. De exemplu : VVC— (*oaspete*), CCCVV— (*streășină*) etc.

După criteriul lungimii, silabele se repartizează astfel ¹⁸:

Foneme :	1	2	3	4	5
	V(7, 11) ¹⁹	VV (1,5)	VVC (0,18)	CVVV (0,10)	CCVVC (0,18)
	7,11	VC (5,2)		CVVC (0,8)	CCVCC (0,04)
		CV (54,8)	VCC (0,16)	CVCC (1,4)	CCCVC (0,06)
		61,5	CVV (4,7)	CCVV (0,24)	0,28
			CVC (15,8)	CCVC (1,05)	
			CCV (4,7)	CCCV (0,35)	
			25,54	4	

Ca și în cazul cuvintelor monosilabice, silabele cu 4 foneme includ cele mai multe (7) *tipuri* de grupare diferită a fonemelor. *Categoria* silabică cea mai frecventă de grupare diferită a fonemelor este însă aceea de 2 foneme (61,5%), iar cea mai puțin frecventă aceea de 5 foneme într-o silabă (0,28%).

Ca *organizare*, remarcăm că, abstracție făcând de natura fonemelor care alcătuiesc silabele, analiza tipurilor înregistrate pune în evidență în cele mai multe cazuri secvențe simetrice. Secvențele simetrice reprezintă structuri proprii limbii române, indiferent în ce sens sînt citite. Din cele 19 tipuri silabice distincte, 14 (adică 73,6%) au o asemenea natură. Acestea sînt :

I (structuri simetrice care au o singură variantă) V; VV; CVVC; CVC; CCVCC

II (structuri simetrice cu doi termeni) VVC / CVV; VC / CV; VCC / CCV; CVCC / CCVC

Celelalte 5 (26,3%) sînt asimetrice :
CVVV; CCVV; CCVVC; CCCV; CCCVC

¹⁸ Pe baza cifrelor date de P.B. Dencș, *art. cit.*, p. 61, am stabilit următoarele procente pentru lungimea silabelor care compun cuvintele din limba engleză scrisă :

foneme	procente (%)
1	13,3
2	43,8
3	32,5
4	8,7
5	1,4
6	0,1

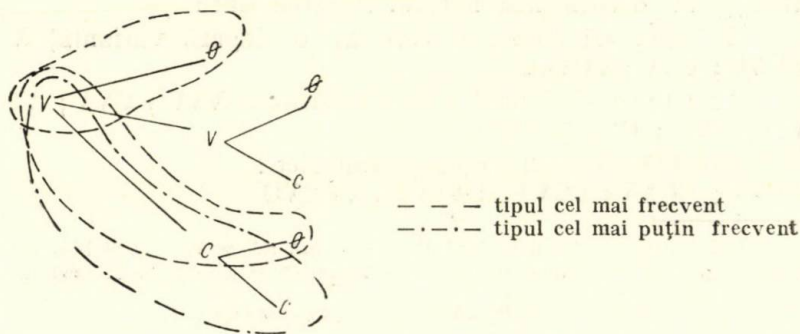
¹⁹ Cifrele din paranteze reprezintă procente.

Preponderența tipurilor silabice simetrice este mai evidentă atunci cînd adăugăm frecvențele lor. Astfel, structurile simetrice cu o singură variantă totalizează 25,25% din silabe; structurile simetrice cu 2 termeni — 72,19%; iar structurile asimetrice numai 0,93%. Este de observat că organizarea simetrică, specifică celor mai multe silabe din cuvintele polisilabice, este caracteristică și silabelor din cuvintele monosilabice.

În ceea ce privește distribuția consoanelor în raport cu vocalele, avem în vedere următoarele: structurile simetrice frecvente pun în evidență silabe echilibrate în care consoanele se distribuie în moduri asemănătoare de o parte sau de alta a nucleului vocalic. *Restricții* în ce privește distribuția fonemelor apar mai ales în cazul tipurilor silabice asimetrice, în sensul următor: o vocală poate fi precedată dar nu și urmată de un segment consonantic complex CCC; o consoană poate fi urmată dar nu și precedată de un segment vocalic complex VVV.

Cîteva mențiuni se referă la începutul și, respectiv, sfîrșitul vocalic sau consonantic al silabelor:

1. a) silabe care încep cu vocală (14, 15 din total) și care pot avea una dintre variantele:

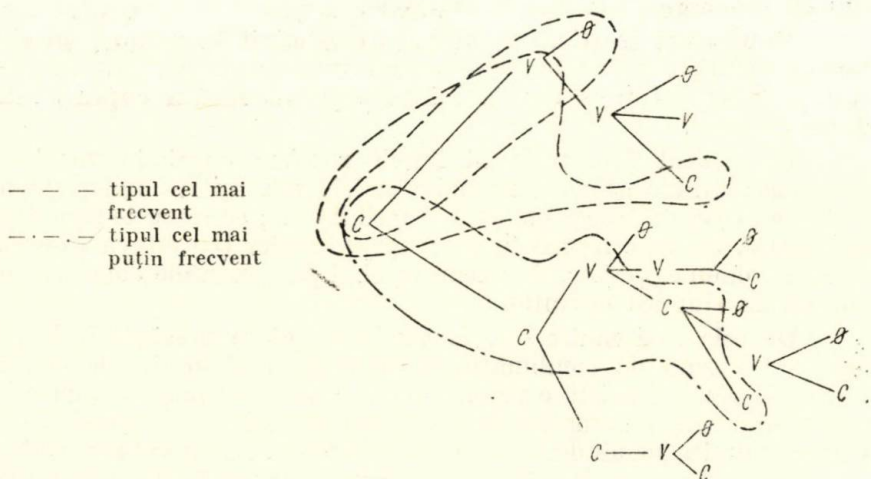


Dintre acestea, cele mai frecvente sînt silabele de tip V (7,11) și VC (5,2), iar cel mai puțin frecvente, silabele de tip VCC (0,16);

b) silabe care încep cu consoană (84,22 din total) și care pot avea una dintre variantele indicate în fig. de la p. 117.

Dintre acestea, cele mai frecvente sînt silabele de tip CV (54,8) și CVC (15,8), iar cel mai puțin frecvente, silabele de tip CCVCC (0,04).

2. a) silabe care se termină în vocală (73,52% din total) dintre care silabele de tip CV sînt cele mai frecvente, iar silabele de tip CVVV — cel mai puțin frecvente.



b) silabe care se termină în consoană (24,93% din total), dintre care silabele de tip CVC sînt cele mai frecvente, iar silabele de tip CVCC — cel mai puțin frecvente.

Din această descriere se observă că tipul silabic cel mai frecvent este CV; cel mai puțin frecvent — CCVVCC. Acest raport între cel mai frecvent / cel mai puțin frecvent tip silabic este *specific* pentru cuvintele polisilabice, întrucît cuvintele monosilabice pun în evidență alte date.

Silabele deschise (terminate în vocală) sînt mai frecvente decît silabele închise ²⁰, fapt care deosebește, de asemenea, cuvintele polisilabice de monosilabe.

În ceea ce privește *aderența*, capacitatea de combinare a acestor silabe între ele, acest aspect este integrat cercetării fonologice la nivelul imediat superior, structura cuvintelor polisilabice.

²⁰ Cf. și Maria Mărdărescu, *Frecvența silabelor deschise și închise în limba română literară*, p. 463—467.

2.2.1.0. Cuvintele

În cele ce urmează ne propunem să cercetăm comportamentul silabelor — material funcțional pe care l-am inventariat pentru stabilirea structurilor silabice cele mai frecvente — în cadrul unor unități fonologice mai mari, cuvintele.

După cum se știe, cuvîntul a fost definit în moduri diferite, fiecare definiție avînd în vedere anumite aspecte ale problemei²¹. Am preferat definirea cuvîntului ca segment minim cuprins între două pauze²².

Inventarul fonematic al limbii române cuprinde vocale și consoane; dintre acestea, numai vocalele pot apărea independent, alcătuiind cuvinte (monosilabice) proprii limbii; consoanele formează, în general, cuvinte numai în prezența vocalelor (cu cîteva excepții, în cazul onomatopeelor, de exemplu, cînd și segmentele consonantice pot forma singure cuvinte).

De cele mai multe ori, fonemele nu ni se prezintă izolat, ci apar sub formă de combinații, în secvențe mai largi, silabele. În structura silabică există o primă condiționare a elementelor, datorită căreia unitățile fonematice se leagă nu numai prin succesiune, ci și prin anumite reguli de dependență reciprocă. La rîndul lor, silabele sînt organizate în cuvinte²³, care cuprind secvențe de silabe succesive și de asemenea interdependente. În alți termeni, lanțurile de foneme alcătuiesc silabe, iar lanțurile de silabe formează cuvinte polisilabice²⁴.

Înlanțuirea fonemelor în silabe, cît și a silabelor în cuvinte, nu se produce numai conform unor reguli de concatenare, ci și conform unor legi statistice; de aceea, secvențele fonematice și, mai departe, secvențele silabice alcătuiesc cuvintele, funcționînd ca evenimente cu probabilități condiționate.

Materialul la care ne referim cuprinde circa 3 760 de cuvinte din texte diverse, aparținînd beletristicii (literatura română a secolului al XX-lea).

²¹ Cf. A. Rosetti, *Remarques sur la définition du „mot”*, p. 261 — 262.

²² Cf. Ch. F. Hockett, *A Course in Modern Linguistics*, p. 621.

²³ O cercetare asupra unităților lexice ale limbii române în *dicționar* stabilește că din 50 000 de cuvinte, cuprinse în DLRM, numai 2 384 sînt monosilabe (Cf. Sanda Golopenția-Eretescu, *art. cit.*, p. 59), deci monosilabele reprezintă 4,7%. Într-o cercetare pe care am întreprins-o asupra lungimii cuvintelor în *texte*, am constatat că frecvența cuvintelor monosilabice românești nu depășește 39,37%. (*Vezi capitolul 2, 2, 2*, p. 138).

²⁴ Pentru unele considerații referitoare la construcția lanțurilor lingvistice, cf. Sorin Stati, *La construction des chaînes linguistiques potentielles*, p. 267 — 280.

În inventarierea tipurilor de cuvinte, nu am luat în considerație numele proprii, prescurtările, cifrele.

Am avut în vedere cuvinte polisilabice de 2—6 silabe (numărul maxim de silabe întâlnit în cuvintele din textele pe care le-am parcurs).

Dintre cuvintele înregistrate, 1 824 — sînt de 2 silabe; 1 110 — de 3 silabe; 748 — de 4 silabe; 74 — de 5 silabe; 9 — de 6 silabe. Exprimat în procente, acest lucru înseamnă :

Cuvinte de	2 silabe	—	48,5%
	3 „	—	29,4
	4 „	—	19,8
	5 „	—	1,9
	6 „	—	0,2

Am urmărit repartizarea diverselor tipuri silabice în cadrul cuvintelor.

2.2.1.1. Cele 1 824 de cuvinte cu 2 silabe pe care le-am înregistrat au următoarele structuri :

2 SILABE

Silaba : I	II	Exemple	%
1	2	3	4
V —	VV	/aia/ (<i>aia</i>)	0,12
	VC	/aur/ (<i>aur</i>)	0,6
	CV	/ușă/ (<i>ușă</i>)	5,34
	CVV	/avea/ (<i>avea</i>)	0,6
	CVVV	/aveau/ (<i>aveau</i>)	0,12
	CVVC	/aveam/ (<i>aveam</i>)	0,24
	CVC	/omul/ (<i>omul</i>)	3,84
	CVCC	/efekt/ (<i>efect</i>)	0,48
	CCV	/afla/ (<i>afla</i>)	0,6
	CCVC	/ekran/ (<i>ecran</i>)	0,36
VV —	CV	/iată/ (<i>iată</i>)	0,72
	CVC	/uitat/ (<i>uitat</i>)	0,6
VVC —	CV	/oaste/ (<i>oaste</i>)	0,12

1	2	3	4
VC —	CV	/unde/ (<i>unde</i>)	3,6
	CVV	/intii/ (<i>intii</i>)	1,77
	CVC	/urlet/ (<i>urlet</i>)	0,96
	CVCC	/invers/ (<i>invers</i>)	0,36
	CCV	/aspru/ (<i>aspru</i>)	1,44
	CCVC	/intrat/ (<i>intrat</i>)	0,24
CV —	V	/bău/ (<i>bău</i>)	0,24
	VV	/tăia/ (<i>tăia</i>)	2,4
	VC	/kite/ (<i>cîte</i>)	0,48
	VCC	/fiind/ (<i>fiind</i>)	0,36
	CV	/masă/ (<i>masă</i>)	22,04
	CVV	/butoi/ (<i>butoi</i>)	3,6
	CVVV	/vedeau/ (<i>vedeau</i>)	0,36
	CVVC	/puteam/ (<i>puteam</i>)	0,48
	CVC	/pilaf/ (<i>pilaf</i>)	10,8
	CVCC	/kurind/ (<i>curînd</i>)	2,28
	CCV	/patru/ (<i>patru</i>)	0,6
	CCVVC	/viklean/ (<i>viclean</i>)	0,12
	CCVC	/kodrul/ (<i>codrul</i>)	0,48
	CCVCC	/kuprins/ (<i>cuprins</i>)	0,12
CVV —	VV	/foaie/ (<i>foaie</i>)	0,12
	CV	/haină/ (<i>haină</i>)	2,4
	CVC	/seamăn/ (<i>seamăn</i>)	0,12
	CCV	/piatră/ (<i>piatră</i>)	0,24
	CCVC	/pietriş/ (<i>pietriş</i>)	0,12
CVVC —	CV	/poartă/ (<i>poartă</i>)	2,04
CVC —	CV	/salva/ (<i>salva</i>)	7,2
	CVV	/luptau/ (<i>luptau</i>)	1,2
	CVC	/singur/ (<i>singur</i>)	3,3
	CVCC	/răknind/ (<i>răcnind</i>)	0,24
	CCV	/simplă/ (<i>simplă</i>)	0,6
	CCVV	/păstrau/ (<i>păstrau</i>)	0,12
	CCVC	/dispreţ/ (<i>dispreţ</i>)	0,24
CCV —	VV	/treia/ (<i>treia</i>)	0,12
	VVC	/graiul/ (<i>graiul</i>)	0,48
	VC	/trăit/ (<i>trăit</i>)	0,12

1	2	3	4
	CV	/stare/ (<i>stare</i>)	3,84
	CVV	/premiu/ (<i>premiu</i>)	0,48
	CVC	/brațul/ (<i>brațul</i>)	2,28
	CVVC	/premiul/ (<i>premiul</i>)	0,12
	CVCC	/privind/ (<i>privind</i>)	0,96
CCVV —	CV	/spaimă/ (<i>spaimă</i>)	0,6
	CVCC	/skuipînd/ (<i>skuipînd</i>)	0,12
CCVVC —	CV	/proaspăt/ (<i>proaspăt</i>)	0,84
CCVC —	CV	/frunte/ (<i>frunte</i>)	1,44
	CVV	/fruntea/ (<i>fruntea</i>)	0,24
	CVC	/plimbat/ (<i>plimbat</i>)	0,84
	CVCC	/stîrnesk/ (<i>stîrnesk</i>)	0,12
CCCV —	CV	/strada/ (<i>strada</i>)	1,44
	CVV	/sklipea/ (<i>sclipea</i>)	0,12
CCCVC —	CV	/strînse/ (<i>strînse</i>)	0,36

Cuvintele cu două silabe se organizează în 64 de tipuri distincte²⁵. Exprimată în număr de foneme, lungimea unui cuvînt disilabic înseamnă: de la 3 foneme (tipurile: V—VV, V—VC, CV—V, etc.), de exemplu: /aia/ (*aia*), /aur/ (*aur*), /bău/ (*bău*), pînă la 8 foneme (tipurile CCVV—CVCC, CCVC—CVCC), de exemplu: /skuipînd/ (*skuipînd*), /stîrnesk/ (*stîrnesk*).

Tipurile silabice cele mai frecvente sînt: CV—CV (de exemplu: /masă/ (*masă*) = 22,04% (adică aproape 1/5 din totalul cuvintelor cu 2 silabe) și CV—CVC, (de exemplu /kovor/ (*covor*)) = 10,8% (adică aproape 1/10 din totalul cuvintelor cu 2 silabe).

²⁵ O comparație cu tipurile de cuvinte disilabice, cele mai frecvente din limbile italiană și engleză, ne arată că toate aceste tipuri există și în limba română (deși nu totdeauna sînt cele mai frecvente). Tipurile din limba italiană la care ne referim sînt: baba (care după cum am văzut este cel mai frecvent și în limba română), babba, bbaba, babaa, baaba. Tipurile din limba engleză sînt: babab, ba ba, (ambele cu o mare frecvență și în limba română), babbab, babba, bbaba. Aceste tipuri au fost înregistrate de Jiří Krámský, *A Quantitative Phonemic Analysis of Italian Mono-, Di-, and Trisyllabic Words*, p. 140. Limba română cunoaște, de asemenea, unele dintre cele mai frecvente structuri de cuvinte existente în limbile germană și eehă: aab, aba, bba, baab, baba, abba. Aceste tipuri au fost înregistrate de Vilém Mathesius, *La structure phonologique du lexique du Tchèque moderne*, p. 168.

În unele cazuri remarcăm construcții simetrice : prima silabă poate fi permutată cu a doua silabă (sau invers), obținându-se cuvinte proprii limbii române. De exemplu : V—CV / CV—V ; VV—CV / CV—VV ; CV—CCV / CCV — CV, etc.

Într-o serie de cuvinte, ambele silabe au structuri identice : CV—CV ; CVC—CVC.

Un caz particular este acela în care un cuvânt are aceeași structură silabică, fie că se citește începînd de la inițială, fie de la finală. De exemplu : VC—CV ; CV—VC ; CVC—CVC.

Specificul acestor cuvinte, reprezentînd structura lor bimembră, este faptul că ele conțin o silabă inițială și o silabă finală, legate direct (spre deosebire de cuvintele cu trei și mai multe silabe, unde apar una sau mai multe silabe mediale).

În legătură cu succesiunea silabelor, observăm următoarele : după cum se știe, prima silabă dintr-un cuvânt poate fi, în general, oricare dintre silabele care fac parte din inventarul de silabe inițiale ale limbii respective (limitările sînt numai de ordin contextual, fiecare cuvânt fiind condiționat de cuvîntul anterior, pentru ca enunțul să aibă un sens).

După silabele inițiale urmează însă anumite restricții. În acest sens, în cazul cuvintelor de două silabe, silabele inițiale care permit după ele apariția celor mai multe tipuri de silabe sînt :

CV — îi pot urma 14 tipuri silabice)

V — („ „ „ 10 „ „)

CCV— („ „ „ 9 „ „)

În ceea ce privește capacitatea de combinare a silabelor finale, silaba de tip —CV este cea mai aderentă (ea poate urma tuturor tipurilor de silabe inițiale pe care le-am înregistrat) ; silaba de tip —CVC poate urma după 8 din cele 13 silabe inițiale care alcătuiesc această categorie de cuvinte. Celelalte silabe finale prezintă indici mai mici de aderență.

Silabele cele mai puțin aderente sînt :

la inițială : CVVC—, CCVVC—, CCCVC—, care nu pot fi urmate, fiecare, decît de un singur tip de silabă finală (și anume de —CV, silaba cea mai aderentă, după cum am văzut mai sus) ;

la finală : —V, —CCVCC, —CCVVC, care nu pot urma, fiecare, decît unui tip de silabă inițială (din nou lui CV—). Este de observat că toate aceste tipuri silabice (cu excepția lui —V) reprezintă struc-

turi complexe, ceea ce explică în mare măsură gradul lor redus de combinabilitate ²⁶.

2.2.1.2. Cele 1 110 cuvinte cu *trei* silabe, au următoarele structuri :

3 SILABE

Silaba : I II		III	Exemple	%
1	2	3	4	5
V-	V-	CV	/audă/ (<i>audă</i>)	0,4
		CVV	/auzea/ (<i>auzea</i>)	0,4
		CVVV	/auzeau/ (<i>auzeau</i>)	0,2
		CVVC	/auzeam/ (<i>auzeam</i>)	0,4
		CVC	/auzit/ (<i>auzit</i>)	0,2
	CV-	V	/odăi/ (<i>odăii</i>)	0,4
		VV	/ăceia/ (<i>aceea</i>)	0,4
		CV	/opusă/ (<i>opusă</i>)	4
		CVV	/atitea/ (<i>atitea</i>)	0,4
		CVVC	/avizier/ (<i>avizier</i>)	0,2
		CVC	/atitor/ (<i>atitor</i>)	1,6
		CVCC	/apăsînd/ (<i>apăsînd</i>)	0,4
		CCV	/asupra/ (<i>asupra</i>)	0,8
	CVV-	CV	/ușoară/ (<i>ușoară</i>)	0,8
		CVC	/aviator/ (<i>aviator</i>)	0,4
	CVC-	CV	/inertă/ (<i>inertă</i>)	2,8
		CVV	/ăcestea/ (<i>acestea</i>)	0,8
		CVC	/ăcestor/ (<i>acestor</i>)	0,4
		CVCC	/odihnesk/ (<i>odihnese</i>)	0,2
	CCV-	CVV	/apropie/ (<i>apropie</i>)	0,8
		CVC	/otrăvit/ (<i>otrăvit</i>)	1,2
	CCVV-	CV	/aproape/ (<i>aproape</i>)	0,4
VV-	CV-	CV	/iubite/ (<i>iubite</i>)	0,8

²⁶ Am putea vorbi, în legătură cu gradele diferite de aderență ale silabelor, de o situație analogă celei din planul fonematic, *atracția* și *repulsia* fiind tendințele de bază care reglementează interrelația dintre sunete în diferite poziții, în cadrul anumitor unități. Cf. pentru analiza unor astfel de raporturi, J. Krámský, *art. cit.*, p. 131.

1	2	3	4	5
		CVV	/ieşirea/ (<i>ieşirea</i>)	0,2
	CVC-	CV	/ailaltă/ (<i>ailaltă</i>)	0,2
VVC-	CV-	CV	/iernile/ (<i>iernile</i>)	0,2
VC-	CV-	VCC	/inkeind/ (<i>încheind</i>)	0,2
		CV	/ordonă/ (<i>ordonă</i>)	3
		CVV	/istoria/ (<i>istoria</i>)	0,4
		CVC	/ultimul/ (<i>ultimul</i>)	2
		CVCC	/infigind/ (<i>înfigind</i>)	0,4
	CVC-	CV	/askunde/ (<i>ascunde</i>)	2
		CVC	/aşteptat/ (<i>aşteptat</i>)	1,4
		CVCC	/înçerkind/ (<i>încercînd</i>)	0,8
		CCV	/albaştri/ (<i>albaştri</i>)	0,2
		CCVC	/întimplat/ (<i>întîmplat</i>)	0,4
	CVVC-	CV	/întoarse/ (<i>întoarse</i>)	0,2
	CCV-	CV	/isprava/ (<i>isprava</i>)	1,6
		CVC	/mpre3ur/ (<i>împrejur</i>)	0,2
	CCVC-	CV	/îmbrînçi/ (<i>îmbrînci</i>)	0,2
		CVC	/îmbrîncit/ (<i>îmbrîncit</i>)	0,2
		CVCC	/îndreptind/ (<i>îndreptînd</i>)	2
VCC-	CV-	CV	/altceva/ (<i>altceva</i>)	0,4
	CCV-	CV	/eksplikă/ (<i>explică</i>)	0,2
CV-	V-	CV	/băuse/ (<i>băuse</i>)	2,2
		CVC	/violet/ (<i>violet</i>)	0,2
	VV-	CV	/kuiele/ (<i>cuiete</i>)	0,2
		CVC	/băiatul/ (<i>băiatul</i>)	0,2
	CV-	V	/gîfii/ (<i>gîfii</i>)	0,4
		VV	/biruie/ (<i>biruie</i>)	0,8
		VVCC	/lokuiesk/ (<i>locuiesc</i>)	0,2
		VC	/sikîit/ (<i>sicîit</i>)	0,6
		VCC	/pirîind/ (<i>pîrîind</i>)	0,2
		CV	/seçeră/ (<i>seceră</i>)	20
		CVV	/gîtului/ (<i>gîtului</i>)	2,8
		CVC	/bukătar/ (<i>bucătar</i>)	6,4
		CVCC	/reçitînd/ (<i>recitînd</i>)	0,4
		CCVC	/binoklul/ (<i>binoclul</i>)	0,2
	CVV-	CV	/răkoare/ (<i>răcoare</i>)	2

1	2	3	4	5
	CVVC-	CV	/găsească/ (<i>găsească</i>)	0,2
	CVC-	CV	/liniște/ (<i>liniște</i>)	1,2
		CVC	/balansul/ (<i>balansul</i>)	1,2
		CCV	/ferestre/ (<i>ferestre</i>)	0,2
	CCV-	CV	/vitregă/ (<i>vitregă</i>)	0,2
		CVC	/petrecut/ (<i>petrecut</i>)	0,2
	CCVC-	CVC	/kuprinsul/ (<i>cuprinsul</i>)	0,2
CVV-	CV-	CV	/hainele/ (<i>hainele</i>)	0,8
CVVC-	CV-	CVC	/pierdusem/ (<i>pierdusem</i>)	0,2
CVC-	CV-	V	/hirtie/ (<i>hirtie</i>)	0,4
		VVC	/războiul/ (<i>războiul</i>)	0,2
		CV	/kărțile/ (<i>cărțile</i>)	8,4
		CVV	/marginea/ (<i>marginea</i>)	0,4
		CVC	/farmekul/ (<i>farmecul</i>)	2,4
		CVCC	/testament/ (<i>testament</i>)	0,2
	CVV-	CV	/dumneata/ (<i>dumneata</i>)	0,6
	CVC-	CV	/răspunse/ (<i>răspunse</i>)	1,2
		CVC	/dezvoltat/ (<i>dezvoltat</i>)	0,6
	CCV-	CV	/dezbrațe/ (<i>dezbrațe</i>)	0,6
	CCVV-	CV	/distrează/ (<i>distrează</i>)	0,2
CVCC-	CV-	VC	/sophtuos/ (<i>sophtuos</i>)	0,2
CCV-	V-	CVC	/vroisem/ (<i>vroisem</i>)	0,2
	VVC-	CVC	/proektul/ (<i>proiectul</i>)	0,2
	CV-	VV	/sfātuia/ (<i>sfātuia</i>)	0,8
		CV	/fratele/ (<i>fratele</i>)	4
		CVV	/drakului/ (<i>dracului</i>)	0,8
		CVC	/prefăcut/ (<i>prefăcut</i>)	0,6
		CVCC	/prevenind/ (<i>prevenind</i>)	0,2
	CVV-	CV	/frumoasă/ (<i>frumoasă</i>)	0,2
	CVC-	CV	/kredință/ (<i>credință</i>)	0,8
		CVC	/prezentat/ (<i>prezentat</i>)	0,4
		CVCC	/studentesk/ (<i>studentesc</i>)	0,4
CCVC-	CV-	CV	/krestele/ (<i>crestele</i>)	0,6
		CVC	/plantonul/ (<i>plantonul</i>)	0,2
CCCV-	CV-	CV	/strigăte/ (<i>strigăte</i>)	0,2

Cuvintele de trei silabe se organizează în 94 de tipuri distincte. Fonematic, lungimea lor este de 4 foneme (tipurile : V—V—CV, V—CV—V), de exemplu : /audă/ (*audă*), /odăi/ (*odăii*) pînă la 10 foneme (tipurile : VC—CCVC—CVCC, CCV—CVC—CVCC), de exemplu : /îndreptînd/ (*îndreptînd*), /studentesk/ (*studentesc*).

Tipurile silabice cele mai frecvente sînt : CV—CV—CV de exemplu : /sečeră/ (*seceră*) = 20% (adică, 1/5 din totalul cuvintelor de 3 silabe) și CVC—CV—CV = 8,4% (de exemplu : /kărțile/ (*cărțile*)).

În cazul acestor silabe, simetria constă în posibilitatea de repetare a unor silabe cu structuri identice la inițială și la finală cuvintului, silaba medială constituind „axa de simetrie”. De exemplu :

V—CV—V ; CV—V—CV ; CV—VV—CV ; CV—CVVC—CV ;
CV—CVC—CV ; CV—CCV—CV ; CVC—CV—CVC.

Cuvintele în care toate trei silabele au structuri identice sînt cele de tipul : CV—CV—CV ; CVC—CVC—CVC. În cazul particular CVC—CVC—CVC structura cuvintelor rămîne aceeași, fie că sînt citite de la inițială spre finală, fie de la finală spre inițială.

În cuvintele cu trei silabe, trecerea de la inițială la finală se face prin intermediul unei silabe mediale.

Dintre silabele inițiale, cele care permit diversitatea maximă a silabelor mediale care le urmează sînt :

CV — (îi pot urma 8 tipuri silabice)

V — (îi pot urma 6 tipuri silabice)

VC— și CCV— (le pot urma cîte 5 tipuri silabice)

Posibilitățile de combinare ale silabelor mediale se stabilesc în funcție de două elemente : a) silabele inițiale, care le precedă și b) silabele finale, care le urmează.

Silabele mediale care pot urma celor mai multe silabe inițiale sînt : —CV— (tipul acesta apare după toate cele 14 tipuri de silabe inițiale înregistrate de noi, în cuvintele trisilabice) : —CVC— (apare după 6 din cele 14 tipuri de silabe inițiale).

Silabele mediale care pot fi urmate de cele mai multe silabe finale sînt :

—CV— (îi pot urma 43 de tipuri silabice)

—CVC— (îi pot urma 19 tipuri silabice)

Cele mai aderente tipuri de silabe finale sînt :

—CV (poate urma după 35 din cele 40 de tipuri de silabe mediale)

—CVC (poate urma după 23 din cele 40 de tipuri de silabe mediale).

Silabele cele mai puțin aderente sînt și în acest caz unități cu o structură complexă :

— la inițială : VVC—, VCV—, CVV—, CVVC—, CVCC—, CCCV—;

— la finală : —CVVV, —VVCC (fiecare dintre aceste tipuri pot fi urmate, respectiv precedate, de cîte un singur tip silabic);

— în poziție medială : —CCVV— (poate fi precedat numai de V— și urmat numai de —CV; —CVVC— (poate fi precedat numai de CCV— și urmat numai de —CVC).

2.2.1.3. Cele 748 de cuvinte cu 4 silabe au următoarele structuri :

4 SILABE

Silaba : I	II	III	IV	Exemple	%
1	2	3	4	5	6
V-	CV-	VC-	CV	/înainte/ (<i>înainte</i>)	3
			CVV	/înainte/ (<i>îainte</i>)	0,5
		CV-	CV	/apukase/ (<i>apucase</i>)	6
			CVV	/apusului/ (<i>apusului</i>)	3,5
			CVC	/ereditar/ (<i>ereditar</i>)	4
			CV	/uniforma/ (<i>uniforma</i>)	0,5
	CVC-		VV	/amîndouă/ (<i>amîndouă</i>)	1
	CVC-	CV-	CV	/amintire/ (<i>amintire</i>)	1,5
			CVV	/înălțimea/ (<i>înălțimea</i>)	1
VV-	CV-	CV-	CVC	/oamenilor/ (<i>oamenilor</i>)	1
VC-	CV-	VV-	CVC	/înduioșat/ (<i>înduioșat</i>)	0,5
		CV-	CV	/arborele/ (<i>arborele</i>)	5,5
			CVC	/începutul/ (<i>începutul</i>)	4
			CVCC	/interesant/ (<i>interesant</i>)	0,5

Silaba : I	II	III	IV	Exemple	%
1	2	3	4	5	6
		CVC-	CCV	/înregistră/ (<i>înregistră</i>)	0,5
		CVVC-	CV	/însoțescă/ (<i>însoțescă</i>)	0,5
	CVC-	CV-	CVV	/înkintarea/ (<i>încîntarea</i>)	1
			CVC	/întorsesem/ (<i>întorsesem</i>)	1,5
		CVV-	CV	/înkonoare/ (<i>înconjoare</i>)	1
		CVVC-	CV	/întîlnească/ (<i>întîlnească</i>)	1
	CCV-	V-	CV	/împreună/ (<i>împreună</i>)	1
		CV-	CV	/înkredere/ (<i>încredere</i>)	0,5
			CVC	/înfrigurat/ (<i>înfrigurat</i>)	0,5
	CCVC-	CV-	CV	/îndreptară/ (<i>îndreptară</i>)	0,5
	CCVCC-	CV-	CV	/instinktivă/ (<i>instinctivă</i>)	0,5
CV-	V-	CV-	CV	/puerile/ (<i>puerile</i>)	2,5
		CCV-	CVC	/geografie/ (<i>geografie</i>)	0,5
	CV-	V-	CV	/silueta/ (<i>silueta</i>)	0,5
			CVCC	/călăuzesk/ (<i>călăuzesc</i>)	0,5
		CV-	VCC	/tălăzuiesk/ (<i>tălăzuiesc</i>)	0,5
			CV	/lovitură/ (<i>lovitură</i>)	17
			CVV	/căminului/ (<i>căminului</i>)	1,5
			CVVV	/căpeteniei/ (<i>căpeteniei</i>)	0,5
			CVC	/pasașerul/ (<i>pasagerul</i>)	6,5
		CVV-	CV	/dimineată/ (<i>dimineată</i>)	0,5
		CVC-	CV	/rekunosku/ (<i>recunosc</i>)	0,5
		CCV-	CVV	/răsuflarea/ (<i>răsuflarea</i>)	0,5
	CVV-	CV-	CV	/răzoarele/ (<i>răzoarele</i>)	1,5
	CVC-	CV-	CV	/părintele/ (<i>părintele</i>)	3
			CVC	/liniștitor/ (<i>liniștitor</i>)	1,5
	CCVC-	CV-	CV	/pătrunsese/ (<i>pătrunsese</i>)	1
CVV-	CCCV-	CV-	CV	/paisprezece/ (<i>paisprezece</i>)	0,5
CVC-	CV-	V-	CVC	/kîmpiilor/ (<i>cîmpiilor</i>)	0,5
		CV-	CV	/furculița/ (<i>furculița</i>)	2
			CVV	/portarului/ (<i>portarului</i>)	0,5
			CVC	/fulgerător/ (<i>fulgerător</i>)	2
		CVC-	CVC	/sentimentul/ (<i>sentimen- tul</i>)	0,5

1	2	3	4	5
	CVV-	CV-	VC	/destăinuit/ (<i>destăinuit</i>) 0,5
	CVC-	CV-	CV	/fantastikă/ (<i>fantastică</i>) 1,5
	CCV-	VC-	CV	/konștiință/ (<i>conștiință</i>) 0,5
CCV-	V-	CV-	CVV	/greutatea/ (<i>greutatea</i>) 0,5
	CV-	CV-	CV	/flăcăările/ (<i>flăcăările</i>) 6
		CVV-	CV	/zgomotoase/ (<i>zgomotoa-</i> <i>se</i>) 0,5
		CVC-	CV	/pregătește/ (<i>pregătește</i>) 0,5
			CVV	/frământarea/ (<i>frământa-</i> <i>rea</i>) 0,5
CCVC-	CV-	V-	CVC	/plescăitul/ (<i>plescăitul</i>) 0,5
CCCV-	CV-	CV-	CV	/strigătele/ (<i>strigătele</i>) 0,5
			CVV	/strălucirea/ (<i>strălucirea</i>) 0,5

Cuvintele de patru silabe se organizează în 58 de tipuri distincte. Lungimea lor este : de la 7 foneme (tipurile : V—CV—VC—CV, CV—V—CV—CV, CV—CV—V—CV), de exemplu : /înainte/ (*înainte*), /vietate/ (*vietate*), /silueta/ (*silueta*) la 11 foneme (tipurile : VC—CVC—CVVC—CV, VC—CCVCC—CV—CV, CVV—CCCV—CV—CV), de exemplu : /întilnească/ (*întilnească*), /instinctivă/ (*instinctivă*), /paisprezece/ (*paisprezece*).

Tipul silabic cel mai frecvent este : CV—CV—CV—CV (de exemplu : /lovitură/ (*lovitură*), = 17% (adică, aproape 1/5 din totalul cuvintelor cu 4 silabe).

În aceste cuvinte n-am putut sesiza structurări simetrice.

Cuvintele în care toate patru silabele au structuri identice sînt de tipul : CV—CV—CV—CV.

În cuvintele cu patru silabe, trecerea de la inițială la finală se face prin intermediul a două silabe, pe care le vom numi : medială₁ și medială₂.

Silabele inițiale care permit, după ele, apariția celor mai multe tipuri de silabe mediale₁ sînt :

VC- și CV- (le pot urma cîte 5 tipuri silabice)
CVC- (îi „ „ „ 4 „ „)

Silabele mediale₁ : a) care pot urma celor mai multe silabe inițiale sînt : -CV- (apare după 8 din cele 9 tipuri de silabe inițiale

posibile în cuvinte de patru silabe); -CVC- (apare după 4 din cele 9 tipuri de silabe inițiale); b) cele care pot fi urmate de cele mai multe tipuri de silabe mediale₂ sînt: -CV- (ii pot urma 8 tipuri silabice) -CVC- (ii pot urma 3 tipuri silabice).

Silabele mediale₂ : a) care pot urma celor mai multe silabe mediale₁ sînt: -CV- (poate urma după 10 tipuri silabice); b) cele care pot fi urmate de cele mai multe tipuri de silabe finale sînt: -CV- (ii pot urma 7 tipuri silabice); -CVC- (ii pot urma 4 tipuri silabice).

Cele mai aderente tipuri de silabe finale sînt: -CV, -CVC (pot urma după cîte 6 tipuri din cele 8 tipuri de silabe mediale₂).

Silabele cele mai puțin aderente, avînd de asemenea de cele mai multe ori structuri complexe, sînt:

- la inițială: VV-, CVV-, CCVC- CCCV-,
- la finală: -VVCC, -CVVV;
- în poziție medială: -CCVC-, -CCVCC-, -CCCV-.

2.2.1.4. Cele 74 de cuvinte de 5 silabe au următoarele structuri:

5 SILABE

Silaba : I	II	III	IV	V	Exemple	%
1	2	3	4	5	6	7
V-	CV-	V-	CV-	CV	/epuizeze/ (epuizeze)	2,70
		CV-	CV-	CV	/agățaseră/ (agățaseră)	10,80
				CVV	/atitudinea/ (atitudinea)	5,40
				CVC	/akoperișul/ (acoperișul)	2,70
VC-	CV-	CV-	V-	CVC	/îngăduitor/ (îngăduitor)	2,70
			CV-	CV	/înkura3ară/ (încurajară)	2,70
				CVV	/înkipuirea/ (închipuirea)	2,70
		CVC-	CV-	CV	/argumentele/ (argumentele)	5,40
				CVV	/înțelepçunea/ (înțelepciunea)	2,70
			V-	CV	/încendiare/ (încendiare)	2,70
	CCV-	CV-	CVV-	CV	/îngrozitoare/ (îngrozitoare)	2,70

Silabele mediale₁ : silabele din tipul -CV- sînt cele mai aderente, atît faţă de iniţiala care le precedă (pot urma după toate 5 tipuri de iniţiale posibile), cît şi faţă de mediala₂ care le urmează (pot preceda toate cele 3 tipuri de mediale₂ posibile).

Silabele mediale₂ : de asemenea, silabele -CV- sînt cele mai aderente (pot urma după 3 din cele 4 mediale₁ posibile şi pot preceda toate cele 4 mediale₃ posibile).

Silabele mediale₃ : silabele mediale -CV- pot urma toate (3) silabele mediale₂ şi preceda toate (5) silabele finale.

Silabele finale cele mai aderente sînt -CV, care pot urma după toate 4 silabele mediale₃ posibile.

Silabele cele mai puţin aderente sînt :

la iniţială : CVC-, CCVC- ;

la finală : -VV ;

în poziţie medială : -CCVV-.

Este de observat că, de astă dată, caracterul neaderent al silabelor nu ţine în primul rînd de complexitatea structurii lor, pentru că intervine un alt element, anume caracterul mai complicat al relaţiilor ce se stabilesc între un număr atît de mare de foneme (componente ale silabelor) ale întregului cuvînt ; e vorba deci, de o complexitate la nivelul cuvîntului.

2.2.1.5. Cele 8 cuvinte cu 6 silabe au următoarele structuri :

6 SILABE

S i l a b a						Exemple	%
I	II	III	IV	V	VI		
VC-	CV-	CV-	CV-	CV-	CV	/înşumătăţise/ (înjumătăţise)	33
CV-	CV-	V-	CV-	CV-	CV	/fizioloğică/ (fiziologică)	33
CVC-	CVC-	CV-	CVC-	CV-	CVC	/bombardamentelor/ (bombardamentelor)	33

Şi în cazurile de mai sus — trei tipuri de cuvinte — silabele CV sînt cele mai aderente, în toate poziţiile, ele fiind în acelaşi timp şi singurele structuri care pot apărea în toate aceste poziţii, în orga-

nizarea cuvintelor de 6 silabe. De altfel, observăm că în afară de silabele CV, numai silabele CVC, VC și V mai pot intra în compoziția acestor cuvinte. (Amintim că silabele de tip CV, CVC, V sînt totodată structurile silabice cele mai frecvente din limba română).

2.2.1.5. Cercetarea fiecărui tip de cuvinte, ca și a fiecărei categorii (de 2, respectiv 3, 4, 5, 6 silabe) în parte, este determinată de anumite particularități ale diverselor structuri. Un tip de cuvinte nu „decurge” cu necesitate din altul : cuvintele de 3 silabe nu înseamnă totdeauna = structura unui cuvînt de 2 silabe + structura unui cuvînt de 1 silabă, nici cele de 4 silabe nu se deduc cu necesitate din : un cuvînt de 3 silabe + 1 silabă, etc.

Pentru fiecare tip există modalități specifice de combinare a silabelor : o silabă sau mai multe silabe condiționează pe cea sau pe cele care o sau le succedă. Există secvențe silabice posibile numai într-o anumită categorie de cuvinte, după cum există și secvențe comune mai multor categorii. Ele se deosebesc prin structura, ordinea și frecvența elementelor care le compun.

Analizînd, așa cum ne-am propus, secvențele silabice ca lanțuri lingvistice, observăm că numai unele dintre ele sînt *extensibile*, iar altele nu. Astfel două silabe de tipul V-VV (*aia*) pot fi întrebuintate ca unități superioare, alcătuiind singure cîte un cuvînt. După ele, în general, nu mai urmează alte silabe, cel puțin în majoritatea cazurilor, după cum indică structurile lexicale cele mai frecvente. La fel, trei silabe de tipul V-V-CV (*audă*), patru silabe de tipul VV-CV-CV-CVC (*oamenilor*) etc. Atunci cînd se face *alegerea* unităților care urmează într-un lanț, cunoașterea probabilităților de concatenare a silabelor determină structurarea într-un fel sau altul a secvențelor. Astfel, de exemplu, după o secvență -V-CV-, căreia i-ar putea succeda, la alegere, 8 tipuri silabice diferite, una dintre silabele VV, VVC, CV, CVV, CVVC, CVC, CVCC, CCV, cel mai probabil este să urmeze -CV (0,055) și cel mai puțin probabil -VVC, -VC, -CVVC (0,0006).

Organizarea structurilor silabice cele mai frecvente de 2, 3, 4, 5, 6 silabe se reflectă în tabelul următor ²⁷ :

²⁷ Tabelul cuprinde o descriere generală a secvențelor inventariate, în sensul că se face abstracție de formele lexicale în care apar. Aceeași secvență de 2 silabe, de exemplu V-CV, poate apărea și în cuvinte de 2 silabe, dar și de 3, 4, 5 silabe; aceeași secvență de 3 silabe, de exemplu CVC-CVC-CV, poate apărea nu numai în cuvintele de 3 silabe, ci și în cuvintele de 4 și 6 silabe etc.

SECVENȚE DE 2 SILABE

V	-V	0,0042	VC	-CV	0,062			
	VV	0,0006		CVV	0,010			
	VC	0,003		CVVC	0,0024			
	CV	0,10		CVC	0,025			
	CVV	0,0097		CVCC	0,0048			
	CVVV	0,0012		CCV	0,016			
	CVVC	0,0018		CCVC	0,011			
	CVC	0,036		CCVCC	0,0006			
	CCV	0,011						
	CCVV	0,024						
VV	CCVC	0,0006	VCC	-CV	0,0006			
	-CV	0,0090		CCV	0,0006			
VVC	CVC	0,0084						
	-CV	0,0006						
CVC	CVC	0,0006	CVC	-VV	0,0012	CCVV	-CV	0,004
	-V	0,024		CV	0,12		CVCC	0,0006
CV	VV	0,025		CVV	0,015			
	VVC	0,0018		CVVC	0,0006			
	VVCC	0,0012		CVC	0,039	CCVC	-CV	0,018
	VC	0,0097		CVC	0,005		CVV	0,0012
	VCC	0,003		CCV	0,0042		CVC	0,006
	CV	0,53		CCVV	0,0012		CVCC	0,007
	CVVV	0,055		CCVC	0,003			
	CVVC	0,002						
	CVV	0,042						
	CVC	0,14						
	CVCC	0,018	CVCC	-CV	0,0006	CCVCC -CV 0,0006		
	CCV	0,09						
	CCVVC	0,0006						
	CCVC	0,0042	CCV	-V	0,0012	CCCV	-CV	0,010
	CCVCC	0,0006		VV	0,0006		CVV	0,0006
	CCCV	0,0006		VVC	0,003			
		VVCC		0,0006				
CVV	-CV	0,034		VC	0,0012			
	CVC	0,0012		CV	0,061			
				CVV	0,006			
CVVC	-CV	0,016		CVVC	0,0006			
				CVC	0,022			
				CVCC	0,0048			

SECVENTE DE 3 SILABE

V -V	-CV	0,0012	VV -CV	-CV	0,0036
V -V	-CVV	0,0012	VV -CV	-CVV	0,0006
V -V	-CVVV	0,0006	VV -CVC	-CV	0,0006
V -V	-CVVC	0,0006			
V -V	-CVC	0,0012	VVC -CV	-CV	0,0006
V -CV	-V	0,0012			
V -CV	-VV	0,0006	VC -CV	-VCC	0,0006
V -CV	-VVC	0,0006	VC -CV	-CV	0,023
V -CV	-CV	0,055	VC -CV	-CVV	0,0012
V -CV	-CVV	0,0024	VC -CV	-CVC	0,0097
V -CV	-CVVC	0,0006	VC -CV	-CVCC	0,0012
V -CV	-CVC	0,0048	VC -CVC	-CV	0,014
V -CV	-CVCC	0,0012	VC -CVC	-CVC	0,0042
V -CV	-CCV	0,0024	VC -CVC	-CVCC	0,0024
V -CVV	-CV	0,0024	VC -CVC	-CCV	0,0006
V -CVV	-CVC	0,0012	VC -CVC	-CCVC	0,0012
V -CVC	-CV	0,010	VC -CVVC	-CV	0,0024
V -CVC	-CVV	0,0012	VC -CCV	-CV	0,007
V -CVC	-CVC	0,0012	VC -CCV	-CVC	0,0006
V -CVC	-CVCC	0,0006	VC -CCVC	-CV	0,0012
V -CCV	-CVV	0,0024	VC -CCVC	-CVC	0,0006
V -CCV	-CVC	0,0036	VC -CCVC	-CVCC	0,006
V -CCVV	-CV	0,0012			
			VCV -CV	-CV	0,0006
			VCC -CV	-CV	0,0012
			VCC -CCV	-CV	0,0006
CV -V	-CV	0,013	CVC -CV	-V	0,003
CV -V	-CVC	0,0024	CVC -CV	-VVC	0,0006
CV -VV	-CV	0,0006	CVC -CV	-CV	0,04
CV -VV	-CVC	0,0012	CVC -CV	-CVV	0,0042
CV -CV	-V	0,006	CVC -CV	-CVC	0,011
CV -CV	-VV	0,003	CVC -CV	-CVCC	0,0006
CV -CV	-VVC	0,0006	CVC -CVV	-CV	0,0036
CV -CV	-VVCC	0,0012	CVC -CVC	-CV	0,006
CV -CV	-VC	0,0018	CVC -CVC	-CVC	0,0018
CV -CV	-VCC	0,0006	CVC -CCV	-CV	0,0018
CV -CV	-CV	0,15	CVC -CCVV	-CV	0,0006
CV -CV	-CVV	0,02			

CV	-CV	-CVC	0,022	CVCC	-CV	-VC	0,0006
CV	-CV	-CVCC	0,0012				
CV	-CV	-CCVC	0,0006	CCV	-V	-CVC	0,0006
CV	-CVV	-CV	0,011	CCV	-VVC	-CVC	0,0006
CV	-CVVC	-CV	0,0006	CCV	-CV	-VV	0,0024
CV	-CVC	-CV	0,018	CCV	-CV	-CV	0,02
CV	-CVC	-CVC	0,016	CCV	-CV	-CVV	0,003
CV	-CVC	-CCV	0,0012	CCV	-CV	-CVC	0,0036
CV	-CCV	-CV	0,0006	CCV	-CV	-CVCC	0,0006
CV	-CVC	-CVC	0,0012	CCV	-CVV	-CV	0,0006
CV	-CCVC	-CVC	0,0006	CCV	-CVC	-CV	0,0018
				CCV	-CV	-CVC	0,0024
CVV	-CV	-CV	0,0042	CCV	-CV	-CVCC	0,0012
				CCV	-CV	-CV	0,0024
CVVC	-CV	-CVC	0,0006	CCVC	-CV	-CV	0,0042
				CCV	-CV	-CVC	0,0006

SECVENȚE DE 4 SILABE

V	-CV	-CV	-CV	0,003
V	-CV	-VC	-CVV	0,0006
V	-CV	-CV	-CV	0,0084
V	-CV	-CV	-CVV	0,0042
V	-CV	-CV	-CVC	0,0048
V	-CV	-CVC	-VV	0,0012
V	-CV	-CVC	-CV	0,0006
V	-CVC	-CV	-CV	0,0018
V	-CVC	-CV	-CVV	0,0006
VV	-CV	-CV	-CVC	0,0006
VC	-CV	-VV	-CVC	0,0006
VC	-CV	-CV	-CV	0,006
VC	-CV	-CV	-CVC	0,0048
VC	-CV	-CV	-CVCC	0,0006
VC	-CV	-CVC	-CCV	0,0006
VC	-CV	-CVVC	-CV	0,0006
VC	-CVC	-CV	-CVV	0,0012
VC	-CVC	-CV	-CVC	0,0012
VC	-CVC	-CVV	-CV	0,0012
VC	-CVC	-CVVC	-CV	0,0006

VC-CCV	-V	-CV	0,0012	
VC-CCV	-CV	-CV	0,0006	
VC-CCV	-CV	-CVC	0,0012	
VC-CCVC	-CV	-CV	0,0006	
VC-CCVCC	-CV	-CV	0,0006	
CV-V	-CV	-CV	0,0048	
CV-V	-CCV	-CVC	0,0012	
CV-CV	-V	-CV	0,0024	
CV-CV	-V	-CVCC	0,0006	
CV-CV	-CV	-VVCC	0,0006	
CV-CV	-CV	-CV	0,026	CCV -V -CV -CVV 0,0006
CV-CV	-CV	-CVV	0,0048	CCV -CV-CV -CV 0,006
CV-CV	-CV	-CVVV	0,0006	CCV -CV-CVV-CV 0,0012
CV-CV	-CV	-CVC	0,0090	CCV -CV-CVC-CV 0,0012
CV-CV	-CVV	-CV	0,0006	CCV -CV-CVC-CVV 0,0006
CV-CV	-CVC	-CV	0,0006	
CV-CV	-CCV	-CVV	0,0006	CCVC-CV-V -CVC 0,0006
CV-CVV	-CV	-CV	0,0018	
CV-CVC	-CV	-CV	0,0054	CCCV-CV-CV -CV 0,0006
CV-CVC	-CV	-CVC	0,0024	CCCV-CV-CV -CVV 0,0006
CV-CCV	-CV	-CV	0,0006	
CVV-CCCV	-CV	-CV	0,0006	
CVC-CV	-V	-CVC	0,0006	
CVC-CV	-CV	-CV	0,0018	
CVC-CV	-CV	-CVV	0,0012	
CVC-CV	-CV	-CVC	0,0024	
CVC-CV	-CVC	-CVC	0,0006	
CVC-CVV	-CV	-VC	0,0006	
CVC-CVC	-CV	-CV	0,0018	
CVC-CCV	-VC	-CV	0,0006	

SECVENTE DE 5 SILABE

V	-CV	-V	-CV	-CV	0,0006
V	-CV	-CV	-CV	-CV	0,0024
V	-CV	-CV	-CV	-CVV	0,0006
V	-CV	-CV	-CV	-CVC	0,0012

VC	-CV	-CV	-V	-CVC	0,0006
VC	-CV	-CV	-CV	-CV	0,0006
VC	-CV	-CV	-CV	-CVV	0,0006
VC	-CV	-CVC	-CV	-CV	0,0012
VC	-CV	-CVC	-CV	-CVV	0,0006
VC	-CVC	-CV	-V	-CV	0,0006
VC	-CCV	-CV	-CVV	-CV	0,0006
VC	-CCV	-CV	-CVC	-CV	0,0006
VC	-CCVV	-CVC	-CV	-CV	0,0006
VC	-CCVV	-CVC	-CV	-CVC	0,0006
CV	-V	-CV	-CV	-VV	0,0006
CV	-CV	-V	-CV	-CV	0,0012
CV	-CV	-V	-CVV	-CV	0,0006
CV	-CV	-CV	-CV	-V	0,0006
CV	-CV	-CV	-CV	-CV	0,0042
CV	-CV	-CV	-CV	-CVV	0,0012
CV	-CV	-CV	-CV	-CVC	0,0006
CV	-CVC	-CV	-CV	-CVV	0,0006
CV	-CVC	-CV	-CVV	-CV	0,0012
CVC	-CV	-CV	-V	-CV	0,0006
CCVC	-CV	-CV	-CV	-CVC	0,0006

SECVENȚE DE 6 SILABE

CV -CV -V -CV -CV-CV	0,0018
CV -CV -CV-CV -CV-CV	0,0018
CVC-CVC-CV-CVC-CV-CVC	0,0018

2.2.2. Știind că lungimea cuvintelor pune probleme specifice diferite de la un stil la altul, am efectuat o cercetare aparte, în care am inclus cuvinte din mai multe stiluri ²⁸. Am avut în vedere de astă dată, pe lângă concatenarea diverselor *tipuri* de silabe, lungimea cuvîntului în ansamblul ei. Atenția noastră s-a oprit, astfel, la factorii care determină variații de lungime a cuvintelor.

²⁸ Este de observat că unele cifre vor diferi de cele obținute de noi în cercetarea componentei silabice a cuvintelor (*vezi* 2.2.1.1. — 2.2.1.5., p. 119 ș.u.). În acest caz, însă, avem în vedere și cuvintele monosilabice, după cum includem și cuvinte din alte stiluri (publicistic și științific), care prezintă un număr mai mare de cuvinte lungi. Ceea ce apropie însă cele două serii de rezultate este scăderea probabilității de apariție de la cuvintele de 2 silabe, la cele de 3 silabe etc. pînă la cele de 6 silabe, deci, aceeași ordonare a frecvențelor în raport cu lungimea cuvintelor.

În limba română am înregistrat cuvinte de 0, 1, 2, 3, 4, 5, 6 și 7 silabe. În textele parcurse de noi, frecvența acestora este — după cum s-a văzut — următoarea :

Număr de silabe	S t i l u l			Total %
	beletristic	științific	publicistic	
0	0,64	0,26	0,02	0,96
1	16,48	10,03	12,86	39,37
2	9,74	6,18	8,19	24,11
3	7,08	4,75	6,57	18,40
4	2,84	3,90	3,85	10,59
5	0,55	1,76	1,40	3,71
6	0,14	0,36	0,34	0,84
7	0,02	0,08	0,06	0,16

Din acest tabel rezultă următoarele :

Atît în ansamblu, cît și pentru fiecare tip de texte în parte, *frecvența* cea mai mare o au cuvintele scurte (mono- și bisilabice); cuvintele de 3 și de 4 silabe au o frecvență medie, iar cele de 6 și 7 silabe sînt slab reprezentate. O problemă specială pun cuvintele pe care le-am numit zero-silabice²⁹, pentru că nu alcătuiesc ele însele silabe (acestea sînt cuvinte care se rostesc legate de altele, de exemplu: *-și* din *nu-și*, *n-* din *n-au*, *-l* din *care-l*, *l-* din *l-au* etc.). Cuvintele de acest tip au o frecvență foarte mică și prin aceasta situația lor se apropie de aceea a cuvintelor lungi.

Menționăm că cercetarea noastră, care se referă la texte diverse, alcătuite din cuvinte care aparțin diferitelor compartimente ale vocabularului, ne-a dus la rezultate care se deosebesc de datele care caracterizează lungimea cuvintelor din fondul principal de cuvinte. În fondul principal, cele mai frecvente sînt cuvintele de 2 silabe, după care urmează, cu un număr aproape egal de apariții, cuvintele mono- și trisilabice. Numeric, raportul dintre aceste trei

²⁹ În lucrările consultate de noi, în care se fac referințe la lungimea cuvintelor, n-am întîlnit nici o mențiune pentru această categorie, ceea ce ne face să credem că s-a procedat la includerea unor cuvinte de tip zero-silabic în categorii de cuvinte silabice din care s-a considerat că provin cuvintele respective (cuvinte zero-silabice există și în alte limbi, de exemplu: fr. *l-* din *l'on*, *qu-* din *qu'il*, *n-* din *n'en*, *d-* din *d'une*; germ. *-s* din *geht's*; engl. *-t* din *can't* etc.). Avînd în vedere însă definiția cuvîntului de la care am pornit, cercetarea noastră urmărește aceste cuvinte ca un tip aparte.

categorii de cuvinte este următorul : cuvinte monosilabice = 338, cuvinte bisilabice = 701, cuvinte trisilabice = 348 ³⁰.

Compararea datelor din tabel indică o repartitie în cea mai mare parte asemănătoare pentru toate textele, a cuvintelor de diferite lungimi ³¹. Remarcăm, totuși, câteva deosebiri : între cuvintele de 1 și 2 silabe diferența de frecvență este mai mare în stilul beletristic decît în stilul publicistic și în stilul științific (6,74 în stilul beletristic față de 4,65 în stilul publicistic și, respectiv, 4,67 în stilul științific ; același lucru între cuvintele de 3 și de 4 silabe (4,24 față de 2,72 și, respectiv, 0,85). Observăm, de asemenea, existența unui număr mai mic de cuvinte lungi în textele beletristice, în raport cu textele din celelalte stiluri : termenii tehnici, mai frecvenți în știință și în presă, dau un procent mai mare de cuvinte lungi.

Analizînd repartitia cuvintelor de diferite lungimi, am observat prezența celor de 1—5 silabe în toate textele. Cuvintele cu o frecvență mai mică, cele zero-silabice și mai des cele de 6, 7 și 8 silabe, nu apar într-o serie de texte. În general, cuvintele de acest fel înregistrate de noi se întîlnesc mai ales în cîte un singur text ; uneori le întîlnim repetate de mai multe ori în același text : /amendamentelor, apukăturile, gospodăriile, grohotișurile, radioreportașe, viețuitoarele/ (*amendamentelor, apucăturile, gospodăriile, grohotișurile, radioreportașe, viețuitoarele*) — de cîte două ori ; /gospodăriilor/ (*gospodăriilor*) — de trei ori ; alteleori, același cuvînt apare în mai multe texte /asemănătoare/ (*asemănătoare*) — în două texte. Celelalte cuvinte apar cîte o singură dată, deși fac parte din fondul comun al limbii / individualitate, întrebuintare, preocupările, regularitate, terminologia/ (*individualitate, întrebuintare, preocupările, regularitate, terminologia*) etc. Alte cuvinte sînt rare, specializate în anumite domenii : /feromagnetismul, pericelulare, perivaskulare/ (*feromagnetismul, pericelulare, perivaskulare*), iar altele nu sînt atestate nici în dicționare : /descarașalizării, diamagnetismul, nekonsumptibile/ (*descarașalizării, diamagnetismul, nekonsumptibile*).

Repartitia cuvintelor lungi în texte punînd probleme speciale, am efectuat o comparație cu situația din dicționar ³², pe

³⁰ Cf. Al. Graur, *Încercare asupra fondului principal lexical al limbii române*, p. 157—161. Listele din această lucrare cuprind numai formele de bază ale cuvintelor.

³¹ În cercetarea noastră, bazată numai pe trei stiluri, nu se observă o mare variație de la un stil la altul, așa cum s-a relevat în studiul lungimii cuvintelor pentru alte limbi. Vezi G. Herdan, *Type-Token Mathematics*, p. 118.

³² *Dicționarul limbii române literare contemporane*, Ed. Academiei R.P.R., I, 1955.

baza unui sondaj pe circa 14 000 de cuvinte. Rezultatele au fost următoarele :

În total, cuvintele de peste 6 silabe constituie circa 2,3% din materialul excerptat, față de circa 1%, cît reprezentau în textele parcurse. Din aceasta rezultă : frecvența mai mică a cuvintelor mai lungi ; repartiția și mai mică a acestor cuvinte, existente în limbă (și atestate în dicționar, în forma de bază), în texte.

Nr. de silabe	Total cuvinte %
6	1,89
7	0,35
8	0,10
9	0,007 ³³

În cercetarea asupra lungimii cuvintelor, ne-am propus să stabilim și care sînt cuvintele lungi din limba română ³⁴.

Urmărind această problemă din punct de vedere etimologic, observăm că cele mai multe cuvinte lungi sînt cele noi ³⁵.

Singurele excepții pe care le putem menționa se referă la două cuvinte înregistrate în dicționar : /anastasimatar/ (*anastasi-matar*) < ngr. *anastasimon* + *at* + *ar* ; /blagoveștenie/ (*blagoveștenie*) < v. sl. *blagovesteniĵe* ; celelalte sînt neologisme sau formații recente din limba română. Dintre aceste din urmă, menționăm, desigur, mai ales cuvintele compuse : /antemăsurătoare/ (*antemăsurătoare*), /bine-mirositoare/ (*binemirositoare*) etc.

Compararea cuvintelor lungi din dicționar cu cele din textele parcurse pune în evidență faptul că, în texte, lungirea cuvintelor se datorește mai puțin compunerii (în cazuri ca : /agrozootehnika, feromagnetismul, radioreporta3e/ (*agrozootehnica, feromagnetismul, radioreportaje*)) și mai mult flexiunii. Flexiunea cuvintelor în limba română făcîndu-se cu ajutorul desinențelor (chiar atunci cînd fuziunii desinențelor la diversele teme i se adaugă și alternanțele fonetice) ³⁶, înregistrăm o serie de cazuri cînd cuvintele își măresc corpul fonetic în cadrul flexionării. Iată cîteva exemple din textele parcurse de noi (cifrele reprezintă numărul de silabe ; în paranteze dăm numărul de silabe al formelor de bază (= cuvinte—titlu) ale cuvintelor respective) :

³³ Un singur cuvînt — compus — de 9 silabe : *anatomo-patologică*.

³⁴ În această lucrare ne referim în special la cuvintele mai rare în limbă, cele lungi. Discuții mai ample asupra cuvintelor scurte, de 1, 2 sau 3 silabe, s-au făcut de Al. Graur, *lucr. cit.*, p. 157—161.

³⁵ Vezi și Al. Rosetti, *Teoria vocabularului*, p. 107.

³⁶ *Gramatica limbii române*, I, p. 67.

/automatizate/ (*automatizate*) — 7 (6), /basoreliefurile/ (*basoreliefurile*) — 7 (4), /karakteristiçe/ (*caracteristice*) — 6 (5), /konsiderentele/ (*considerentele*) — 6 (4), /kumpărătorului/ (*cumpărătorului*) — 6 (4), /despăduririle/ (*despăduririle*) — 6 (5), /gospodăriile/ (*gospodăriile*) — 6 (5), /gospodăriilor/ (*gospodăriilor*) — 6 (5), /grohotișurile/ (*grohotișurile*) — 6 (3), /întreprinderilor/ (*întreprinderilor*) — 6 (5), /modificările/ (*modificările*) — 6 (5), /organizatoriçe/ (*organizatorice*) — 7 (6), /partikularitățile/ (*particularitățile*) — 8 (7), /pericelulare/ (*pericelulare*) — 6 (5), /perivaskulare/ (*perivaskulare*) — 6 (5)³⁷ etc.

În texte am întâlnit puține cuvinte lungi în forma lor de bază : /deģeneresćență/ (*degenerescență*), /električitate/ (*electricitate*), /individualitate/ (*individualitate*), /întrebuințare/ (*întrebuințare*), /materializare/ (*materializare*), /regularitate/ (*regularitate*) etc. Alături de acestea ar trebui să menționăm cazul cînd, prin flexiune, o serie de cuvinte nu-și măresc corpul fonetic : /kombativității/ (*combativității*) — 6 (6), /sensibilității/ (*sensibilității*) — 6 (6), /universalității/ (*universalității*) — 6 (6). Într-un singur caz, flexiunea a dus la scăderea numărului de silabe : /proporționalității/ (*proporționalității*) — 6 (7).

Lungirea cuvintelor flexibile constituie o particularitate a limbii române, față de alte limbi, în care flexiunea de număr, caz, etc. se poate realiza numai prin adăugarea unor desinențe consonantice (care se pronunță sau nu) și care nu formează silabe (de exemplu : fr. *chat* / *chats*, *enfant* / *enfants*, sp. *casa* / *casas*, *puerta* / *puertas*, engl. *book* / *books*, *tree* / *trees*, etc.).

Avînd în vedere că lungimea cuvintelor diferă și în funcție de importanța lor în frază, existînd cuvinte principale (substantiv, adjectiv, verb) și cuvinte secundare (pronume, articol, conjuncție, prepoziție)³⁸, am efectuat un sondaj pe circa 1 000 de cuvinte (din fiecare dintre cele trei stiluri studiate de noi³⁹), urmărind lungimea care caracterizează diversele părți de vorbire. Rezultatele acestei cercetări sînt înscrise în tabelul următor :

³⁷ Acest fenomen are loc, de fapt, și la nivelul cuvintelor mai scurte : /momentele/ (*momentele*) — 4 (2), /bunului/ (*bunului*) — 3 (1).

³⁸ Vezi Al. Rosetti, *Teoria vocabularului*, p. 114—115 și *Filozofia cuvîntului*, p. 53.

³⁹ Textele parcurse pentru acest sondaj sînt : un text beletristic (Hortensia Papadat-Bengescu, *lucr. cit.*, p. 404), un text științific (C. Irimie, *lucr. cit.* p. 31), un text publicistic (*Scînteia* nr. 6154, an. XXXIII, p. 1, articolul „Luna cărții la sate”). Cifrele din tabel reprezintă procente.

Nr. de silabe Părți de vorcare	0	1	2	3	4	5	6	7
substantiv		2,2	9,2	9,1	6,8	2,7	1	0,2
verb		1,9	3,9	3,7	1,2	0,7	0,1	
adjectiv		1,3	2,2	1,7	3,1	0,5	0,1	
pronume	0,5	4,2	2,2	0,8	0,2			
articol		3,6	1	0,2				
adverb	0,2	4,3	1,4	0,1	0,2			
numeral		0,7	0,5	0,1				
prepoziție	0,2	15,1	0,9	0,1				
conjunție		8,1	0,8	0,1				
interjecție		0,1						

În acest tabel se reflectă următoarele : cuvintele scurte, zero-silabice, sînt numai cuvintele secundare ; monosilabice, sînt de circa 10 ori mai multe cuvinte secundare monosilabice decît cuvinte principale (4,4 din 44,5) ; cuvintele lungi, de 4 și mai multe silabe, sînt aproape în totalitate cuvinte principale (cu excepția a 0,4 cuvinte secundare de 4 silabe).

Diferențele de la un stil la altul nu sînt semnificative pentru discuția noastră.

Pentru a putea caracteriza limba română în ceea ce privește lungimea cuvintelor, este necesară o comparație cu alte limbi. Dar întrucît nu dispunem de descrieri statistice complete, din care să putem extrage date pentru această problemă, ne limităm la observația că și limba română, ca și alte limbi, preferă cuvintele scurte.

Datele comparative de care dispunem sînt următoarele ⁴⁰ :

⁴⁰ Rezultatele obținute pentru limbile engleză, germană, rusă, aparțin lui G. Herdan, *Language as Choice and Chance*, p. 177—178, și se bazează pe texte de literatură (beletristică) din limbile respective. Pentru limba germană am adăugat, în paranteze, procentele lui G.K. Zipf, *The psycho-Biology of Language: an Introduction to Dynamic Philology*, p. 23. (De notat, că autorul consideră cuvintele de 7—15 silabe ca însumind 0,22% frecvență.) Pentru limba maghiară, procentele sînt stabilite de I. Fónagy, *Die Silbenzahl der ungarischen Wörter in der Rede*, p. 81 (am reținut cifrele pentru limba scrisă, pentru comparație cu cele din limba vorbită, care constituie, de altfel, obiectul lucrării). În paranteze am adăugat procentele stabilite de Edit Vértess, *Statistische Untersuchungen über den Phonetischen Aufbau der ungarischen Sprache*, p. 155. Autoarea notează pentru cuvintele de 7, 8, 9 silabe, respectiv, procentele : 0—0,2 ; 0—0,05 ; 0—0,05. (Cifrele apar în raportul „de la ... pînă la”, întrucît autoarea nu și-a propus să caracterizeze un anumit tip de texte, ci a înregistrat caracteristicile diferiților autori, fapt care justifică asemenea variații.)

Nr. de silabe	L i m b a			
	engleză	germană	rusă	maghiară
1	69,6	<u>48,8</u> (49,76)	33,1	<u>36,36</u> (35—43)
2	18,2	<u>34,2</u> (28,94)	31,1	<u>29,54</u> (27—31)
3	8	<u>12,1</u> (12,93)	21,4	<u>23,86</u> (16—21)
4	3,2	<u>4,1</u> (5,93)	10	<u>8</u> (7—12)
5	0,8	<u>0,8</u> (1,72)	3,1	<u>2,27</u> (1,7—3,5)
6	0,2	(0,50)	0,9	<u> </u> (0,3—1,3)
7	—		0,4	<u> </u> (0—0,2)

Lungimea medie a cuvintelor (ca număr de silabe) în limba engleză americană este de 2,194⁴¹.

Procentele care disting cuvintele cu un anumit număr de silabe de cuvintele cu un alt număr de silabe diferă de la o limbă la alta; ierarhizarea cuvintelor însă, de la cele monosilabice — de rang 1 — până la cele de 7, 8 sau 9 silabe, rămâne aceeași, chiar și pentru limbile în care compunerea ocupă un loc mai important decât în limba română.

⁴¹ Cf. A. Hood-Roberts, *A Statistical Linguistics Analysis of American English*, p. 44. Cu observația autorului: „This table — — — shows clearly the inherent economy in the language, i.e. the most frequently used words are the shortest”. Dată fiind importanța problemei, diferite lucrări de teoria informației fac mențiuni asemănătoare, cu privire la lungimea cuvintelor. Cf. de exemplu, G. Herdan, *lucr. cit.*, p. 87—89, G.A. Miller, *Language and Communication*. O serie de corespondențe pentru diferite limbi sînt stabilite de Edmond Nicolau, *Language et stratégie*, p. 153—177.

3. ENTROPIA

3.0. Studiarea limbii ca mesaj conduce la calcularea cantității de informație conținută într-un mesaj.

Cantitatea de informație se stabilește pe baza probabilităților de realizare a evenimentului studiat, probabilități care se obțin statistic.

Aspectul cantitativ al informației reprezintă de fapt structura statistică a informației. Forma standard în care ni se prezintă informațiile fiind un șir de simboluri, ceea ce interesează este tocmai frecvența simbolurilor (în cazul limbii: foneme, silabe etc.) și diferențele lor posibilități de combinare.

Măsura cantității de informație, dată de evenimentele care se succed cu anumite probabilități (p_i), este *entropia* (H), introdusă de Shannon — și definită prin relația

$$H = - \sum_1^N p_i \log. p_i$$

Noțiunea de entropie este destul de bine cunoscută, astfel încît nu vom insista asupra ei. Reținem mai curînd cîteva relații.

Ca și entropia din termodinamică, entropia din teoria informației măsoară gradul de nedeterminare a fenomenelor cu care lucrăm.

Deși, în comparație cu noțiunea din termodinamică, entropia lingvistică reprezintă o entropie negativă, este util să utilizăm termenul *entropie* ca sinonim pentru *cantitate de informație*. (În orice caz, entropia H rămîne parametrul caracteristic al teoriei informației).

Din punct de vedere matematic, informația și entropia au aceeași măsură cantitativă; dar ele nu sînt identice. Informația

reprezintă o nedeterminare înlăturată. Sensul variației nedeterminării este opus sensului variației informației¹.

3.1. Pe baza formulei pentru stabilirea entropiei H_1 , s-a calculat entropia fonematică a limbii române, luându-se în considerație frecvențele de apariție ale fonemelor românești.

S-au făcut investigații pe mai multe texte (incluzându-se și considerații în legătură cu entropia în funcție de litere)².

S-a ajuns la valoarea: $H_1 = 4,11$ (entropia/literă).

Este de specificat că, în cazul limbii române, diferențele $H_{(\text{literă})}/H_{(\text{fonem})}$ sînt aproape imperceptibile, dat fiind că pentru relația literă/fonem s-a stabilit caracteristica 1,09³.

De asemenea, a fost elaborată o formulă care „leagă” limbajul oral de limbajul scris, stabilind raportul dintre entropia care se obține pe baza literelor și cea fonematică.

Introducînd indicele $\omega =$ „lungimea medie a fonemelor” (adică tocmai indicele 1,09) — se poate calcula următorul raport:

$$H_{(\text{fonem})} = \omega \cdot H_{(\text{literă})}^4.$$

Pentru limba română, entropia fonematică de ordinul 1 este deci:

$$H_1 = 4,47 \text{ (entropia pe fonem)}$$

Considerăm că este util să prezentăm cîteva date comparative pentru entropia altor limbi (limba scrisă)⁵:

limba	entropia			
	H_0	H_1	H_2	H_3
rusă	5	4,35	3,52	3,01
engl.	4,76	4,03	3,32	3,10 ⁶
germ.	4,70	4,10	—	—
franc.	4,70	3,98	—	—
span.	4,70	4,01	—	—

¹ Silviu Guiașu, Radu Theodorescu, *Matematica și informația*, p. 80.

² Ed. Nicolau, C. Sala, Al. Roceric, *Entropia limbii române*, p. 35—53.

³ M. Grumăzescu, *art. cit.*, p. 1450.

⁴ I.M. Iaglom, R.L. Dobrușin, A.M. Iaglom, *Teoria informației și lingvistica*, p. 38.

⁵ Datele au fost extrase din: E.V. Paducheva, *Information Theory and the Study of Language*, in *Exact Methods in Linguistic Research*, p. 148. Valorile pentru H_2 și H_3 se calculează pe baza combinațiilor de cîte două și trei foneme. S-au avut în vedere numai grupările mediale.

⁶ Menționăm că entropia limbii engleze a fost calculată și astfel: $H_1 = 4,76$ $H_2 = 3,32$ $H_3 = 3,1$. G. Herdan, *Type-Token Mathematics*, p. 171.

Se observă că entropia limbii române — H_1 — nu diferă sensibil de cea stabilită pentru alte limbi. Ea atinge însă valoarea cea mai mare. Ar fi de observat că noi am „transpus” textele scrise în notație fonematică, pe cînd datele din acest tabel au la bază o transcriere în litere.

3.2. Am adăugat la entropia calculată pentru toate textele, în funcție de toate frecvențele fonemelor, considerații privitoare la entropia pentru anumite cazuri, în limba română :

— Astfel, pentru cazul particular al fonemelor inițiale :

$$H_{(\text{inițială})} = 4,56$$

Cantitatea mare de informație care se obține prin considerarea numai a inițialei de cuvînt denotă o anumită claritate. Pentru cel ce receptează un mesaj, e foarte important să poată „descifra” sensul lui general de la început. Or, elemente sugestive în acest sens nu pot fi date de inițială decît într-o limbă în care claritatea este un factor organic. O astfel de trăsătură poate fi considerată caracteristică pentru limba română — în raport cu alte limbi în care sensurile transpar numai în cadrul relațiilor dintre elemente mai largi.

3.3. Cunoscînd structura silabelor din cuvintele polisilabice românești, am calculat entropia în funcție de acest element. Structura unei silabe reprezintă formulele de repartiție a vocalelor și consoanelor în cadrul unor *limite* silabice. În interiorul silabelor, fonemele se comportă ca simboluri dependente și de aceea fiecare silabă a fost considerată ca un întreg.

Cantitatea medie de informație pe care o dă un tip de silabă în limba română este :

$$H_{(\text{silabă})} = 2,6329$$

3.4. Cunoscînd probabilitatea de apariție a secvențelor silabice (2—6 silabe) și știind că entropia obținută prin *considerarea* unei silabe (pe rînd, a fiecărui tip silabic) ca simbol independent este :

$$H_1 = - \sum_{i=1}^n p_{(i)} \log_2 p_{(i)}$$

cînd $p_{(i)}$ = probabilitatea de apariție a unui simbol considerat independent :

$$H_1 = 2,6329$$

putem trece la calcularea entropiei în secvențe silabice ⁷.

Ne vom putea referi, deci, în analiza fonologică a cuvintelor polisilabice, la cantitatea medie de informație pe care ne-o dă o secvență — cu elementele, silabele, condiționate reciproc — de 2, 3, 4, 5, 6 silabe.

$$\text{Astfel, } H^2 = - \sum_{i=1}^n \sum_{j=1}^n p(ji) \log_2 p(ji)$$

unde $p(ji)$ = probabilitatea de apariție a unei anumite perechi de simboluri (în cazul nostru, concatenarea a două silabe)

$$H^2 = 5,2070$$

Utilizînd pe H^2 , calculăm H_2 = entropia condiționată de prim ordin (considerînd dependența statistică numai între 2 simboluri, adică între 2 tipuri silabice)

$$\begin{aligned} H_2 &= H^2 - H_1 \\ H_2 &= 2,5741 \end{aligned}$$

În mod analog, utilizînd pe H_2 și pe H^3 , calculăm H_3 = entropia condiționată de ordinul al doilea (considerînd 3 tipuri silabice dependente).

Calcularea lui H^3 se efectuează astfel :

$$H^3 = - \sum_{i=1}^n p(Bi) \log_2 (Bi)$$

unde : n = numărul diferit de combinații de 3 simboluri (tipuri silabice)

B_i = o anumită combinație de 3 silabe

$$\begin{aligned} \text{Astfel, } H_3 &= 3,3702 \\ H_3 &= H^3 - H_2 \\ H_3 &= 0,7961 \end{aligned}$$

Cunoscînd pe

$$\begin{aligned} H^4 &= 0,9856 \\ H^5 &= 0,2261 \\ H^6 &= 0,0492 \end{aligned}$$

⁷ Această modalitate de calculare a entropiei simbolurilor dependente este descrisă de E.V. Paducheva, *Information Theory and the Study of Language* (în) *Exact Methods in Linguistic Research*, p. 130 ș.u.

putem calcula, după formulele respective, și entropia dată de celelalte secvențe, respectiv :

$$H_4 = 0,1895$$

$$H_5 = 0,0366$$

$$H_6 = 0,0126$$

Rezultatele obținute în cazurile de mai sus sînt în acord cu acea caracteristică potrivit căreia entropia scade cu cît scade numărul de posibilități din care se face alegerea simbolurilor. Intuitiv, se poate stabili că $H_1 \geq H_2$; calculele nu fac decît să precizeze valoarea inegalităților de acest tip.

4. CONCLUZII

4.1. Importanța studiilor matematice se bazează pe exactitatea gândirii și pe corespondența dintre faptele reale din limbă și simbolurile prin care încercăm să le redăm. În cazul nostru, acestea sînt cifre.

Aplicînd metoda cantitativă, sîntem în măsură să stabilim anumite proprietăți cu caracter global, care nu se pot detecta prin alte metode. Totodată, se introduc precizări pentru unele date deja cunoscute. Toate acestea scot în evidență caracterul sistematic al limbii.

De altfel, cuantificînd, nu clarificăm direct, imediat, elementele calitative ale limbii pe care o studiem.

Formalizarea limbii în scopuri cibernetice¹ pune în lumină unele aspecte care nu se pot formaliza. Acestea ne ajută să descoperim proprietăți care deosebesc limba de mașină, orice limbă conținînd un material de o altă complexitate decît mecanismul unei mașini.

În concepția noastră însă, descrierea *multilaterală* a unei limbi nu poate fi lipsită de elementele cantitative.

Se știe, de exemplu, că toate limbile indo-europene au în inventarul lor foneme labiale, dentale etc. Cele mai multe limbi de pe glob — din diverse familii lingvistice — au o serie de vocale.

Aceste elemente — și altele, asupra cărora nu ne oprim! — le aseamănă. Dar studiul cantitativ completează informațiile noastre cu privire la comunitatea mai strînsă, marcată, între limbile române, decît între acestea și limbile germanice, de pildă.

¹ Considerăm că pe lingviști îi interesează studiile cantitative în primul rînd în vederea unei cunoașteri mai exacte a limbii. De aceea, nu intrăm în amănunte și nu facem aprecieri privitoare la aplicațiile practice ale cercetărilor de acest gen (mașini de tradus, dispozitive pentru sintetizarea vorbirii, dispozitive pentru recunoașterea a scrisului cu ajutorul mașinilor electronice etc.)

Pe de altă parte, o asemenea cercetare pune în evidență și ceea ce are *specific* fiecare limbă în parte, raportul dintre elementele ei constitutive.

Desigur, în materie de foneme, însăși stabilirea inventarului fonologic are la bază înregistrarea unui număr mare de variante și selectarea doar a *unor* unități. Dar este preferabilă determinarea *proporției* dintre elementele care alcătuiesc o limbă, față de simpla lor enumerare. Cuantificînd, adăugăm o altă dimensiune sistemului, un sistem în care alăturarea segmentelor în secvențe mai mari este limitată de reguli stricte. Cercetarea combinațiilor posibile între foneme și frecvența acestora relevă tocmai regulile sintagmatice ale sistemului.

S-ar putea ca studii ulterioare, unele efectuate cu calculatoare electronice, pe un material lingvistic mai vast, să modifice întrucîtva sau să completeze rezultatele obținute de noi.

De aceea, lucrarea noastră își propune în primul rînd *ilustrarea* metodei pentru limba română. Scopul ei a fost degajarea unor relații care nu pot fi exprimate decît numeric și despre care avem credința că *sînt caracteristice*, alături de alte elemente, structurii fonologice a limbii noastre.

4.2. Pentru a încheia, vom reține din această descriere, următoarele caracteristici fonostatistice românești :

1. Privitor la tipurile de structuri cantitative :

— Sistemul vocalic românesc este constituit din 7 vocale /a, e, i, ă, î, o, u/ și 20 de consoane /p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, h, Ț, ȅ, ȅ, l, r, m, n/; raportul cantitativ vocale/consoane este : 49,4/49,74.

— Segmentele simple (care cuprind cîte o singură vocală, sau o singură consoană) reprezintă 86,21, față de segmentele complexe care totalizează 12,65.

De aceea, la audiere, limba română nu lasă senzația „aglomerării” de elemente sonore; dimpotrivă, este evidentă structura ei armonioasă, în care vocalele și consoanele se succed proporțional.

Atît în cazul grupurilor vocalice, cît și în cazul grupurilor consonantice, există o preferință pentru structurile simetrice : grupurile vocalice simetrice (94,5) sînt covîrșitor mai frecvente decît grupurile asimetrice (5,5); grupurile consonantice simetrice (66) sînt de 2 ori mai frecvente decît grupurile asimetrice (33).

— În ceea ce privește tranzițiile dintre segmentele simple $V \rightarrow C$, $C \rightarrow V$, mai frecvente sînt trecerile $C + V$, care se află față

de $V + C$ în raportul numeric 60,2/25,6 (tranzițiile al căror prim element este C fiind deci de 2,3 ori mai frecvente decât cele în care elementul prim este V).

— În poziții determinate, frecvența fonemelor variază, în comparație cu frecvența lor globală (stabilită în funcție de totalul ocurențelor lor în limba standard, abstracție făcând de condiționarea pozițională): la inițială, în primele trei ranguri se află consoanele; la finală, în primele trei ranguri se află vocalele. Explicațiile care s-au dat acestui fenomen au fost din domeniul diacroniei, pentru inițialele cuvintelor, și din domeniul sincroniei, pentru finalele de cuvânt.

— Un tabel rezumativ, indică următoarea „ierarhie” a fonemelor românești:

Cele mai frecvente segmente simple (considerate numai în raport cu segmentele simple, în toate pozițiile) sînt: $/e, a, i/$.

Segmentele simple cel mai puțin frecvente (după același criteriu) sînt: $/h, \text{ğ}, \text{ʒ}/$.

Variația valorilor frecvenței — de la maximum la minimum — este următoarea:

pentru vocale: între 10,9 și 3,4 — deci 7,5;

pentru consoane: între 7,6 și 0,13 — deci 6,47.

Și de astă dată diferențele de valori — pe plan vocalic, respectiv consonantic — sînt comparabile.

Cele mai frecvente segmente complexe cu două elemente (considerate numai în raport cu segmentele complexe, în toate pozițiile) sînt:

dintre grupurile vocalice simetrice: $/ea, ui, ie/$;

„ „ „ asimetrice: $/\text{îi}/$;

„ „ „ consonantice simetrice: $/nd, nt/$;

„ „ „ asimetrice: $/st, \text{ʃt}, nt/$.

Cele mai frecvente segmente complexe cu peste două elemente (după criteriile de mai sus) sînt:

dintre grupurile vocalice: $/eau, ioa, uia, eiau/$;

„ „ „ consonantice: $/ntr, str, mpl, mpr; kspl/$.

Nu am făcut nici o mențiune pentru segmentele complexe cu frecvențe mici, deoarece aceasta este o trăsătură comună multora dintre ele, neconstituind deci un factor de individualizare.

Înregistrînd mărimi maxime pentru segmentele menționate mai sus, frecvența ia valori medii pentru o serie oarecare de grupuri

(vocalice și consonantice), apoi scade brusc la cifre mici pentru un număr considerabil de secvențe.

2. Privitor la lungimea secvențelor fonematice :

— În general cele mai frecvente silabe sînt cele constituite din cîte două foneme (61,5 din totalul tipurilor silabice); silabele cele mai puțin frecvente sînt cele care cuprind cîte 5 foneme (0,28). Particularizînd, cele mai frecvente structuri sînt de tipurile : V, VC, CV, CVC; cel mai puțin frecvente sînt de tipurile : VCC, CCVCC.

— La nivelul cuvintelor, frecvența cea mai mare o au cele care cuprind cîte una (39,37) sau două (24,11) silabe.

Dintre cuvintele cu două silabe, cele mai frecvente sînt de tipul

					CV—CV
„	„	„trei	„	„	CV—CV—CV
„	„	„patru	„	„	CV—CV—CV—CV
„	„	„cinci	„	„	CV—CV—CV—CV—CV
„	„	„șase	„	„	CV—CV—CV—CV—CV—CV

3. În ceea ce privește entropia limbii române, reținem următoarele valori :

$$H_{(\text{fonem})} = 4,47$$

$$H_{(\text{silabă})} = 2,63$$

Asemenea date individualizează limba română. În același timp, ele reprezintă rezultate conform cu valorile obținute pentru alte limbi, comparabile în general cu limba română.

5. BIBLIOGRAFIE

5. *Bibliografia* cuprinde lucrările utilizate în vederea alcătuirii fonostatisticii limbii române, fie că aceste lucrări sînt citate sau nu direct în text.

- ACADEMIA REPUBLICII POPULARE ROMÂNE, Institutul de lingvistică din București, *Îndreptar ortografic, ortoepic și de punctuație*, București, f.a.
- ACADEMIA REPUBLICII POPULARE ROMÂNE, *Gramatica limbii române*, București, 1963.
- *Actes du sixième Congrès international des linguistes, Paris juillet, 1948*, Paris, 1949.
- AKHMANOVA O.S., *A propos de la relevance sémiologique des sons du langage*, SCL, XI, 3, 1960, p. 311—316.
- AKHMANOVA O.S., MEL'CHUK I.A., FRUMKINA R.M., PADUCHEVA E.V., *Exact Methods in Linguistic Research*, Berkeley—Los Angeles, 1963 (Traducere din lb. rusă de David G. Hays și Dolores V. Mohr).
- ANDREEV N.D., *Modelarea limbii pe baza structurii ei din punct de vedere al statisticii și al teoriei mulțimilor*, în *Probleme de lingvistică matematică*, București, 1960, p. 15—22.
- ANDREYEV N.D., *Models as a Tool in the Development of Linguistic Theory*, Word, XVIII, 1—2, 1962, p. 186—198.
- ANDREYEV N.D., ZINDER L.R., *On the Notions of the Speech Act, Speech, Speech Probability, and Language*, Linguistics, 4, 1964, p. 5—13.
- APOSTEL L., MANDELBROT B., MORF A., *Logique, langage et théorie de l'information*, Paris, 1957.
- AVRAM ANDREI, *Some Thoughts on the Functional Yield of Phonemic Oppositions*, Linguistics, 5, 1960, p. 40—47.
- AVRAM ANDREI, *Cercelări asupra sonorității în limba română*, București, 1961.
- AVRAM ANDREI, *Asupra tipologiei fonologice cantitative*, SCL, XVI, 3, 1965, p. 323—326.
- BAR-HILLEL JEHOSSUA, *Language and Information*, Jerusalem, 1964.
- BLOMMFIELD LEONARD, *Language*, London, 1961.
- BELEVITCH VITOLD, *Langage des machines et langage humain*, Bruxelles, 1956.

- BOSÁK JÁN, *Frequency of Phonemes and Letters in Slovak and Numerical Expression of Some Phonemic Relations*, Jazykovedný Časopis, XVI, 2, 1965, p. 120—136.
- BURTON N.G., LICKLIDER J.C., *Long-range Constraints in the Statistical Structure of Printed English*, American Journal of Psychology, LXVIII, 4, 1955.
- CARROLL JOHN B., *The Study of Language*, Cambridge, 1961.
- CHOMSKY NOAM, *recenzie la*: Vitold Belevitch, *Langage des machines et langage humain*, Bruxelles, 1956, Language, XXXIV, 1, 1958, p. 99—105.
- CHOMSKY NOAM, *Syntactic Structures*, The Hague, 1963.
- CHERRY COLIN, *On Human Communication*, Massachusetts Inst. of Technology, 1957.
- O'CONNOR J.D., TRIMM J.L.M., *Vowel, Consonant, and Syllable—a Phonological Definition*, Word, IX, 2, 1953, p. 103—123.
- CONSTANTINESCU I., CONDREA S., NICOLAU E., *Teoria informației*, București, 1958.
- DENEŠ P.B., *On the Statistics of Spoken English*, Z. f. Phon., XVII, 1, 1964, p. 51—72.
- DOLBY JAMES L., RESNIKOFF HOWARD L., *On the Structure of Written English Words*, Language, XL, 2, 1964, p. 167—196.
- ELLENBERG S., *Zur Lautstruktur des Wortes*, Z. f. Phon., XVII, 1, 1964, p. 21—24.
- FONÁGY I., *Die Silbenzahl der ungarischen Wörter in der Rede*, Z.f. Phon., XIV, 1, 1961, p. 88—92.
- GNEDENKO B.V., HINCIN A.I., *Introducere elementară în calculul probabilităților*, București, 1954.
- GOLOPENȚIA-ERETESCU SANDA, *La structure phonologique des monosyllabes roumains*, CLTA, III, 1966, p. 59—67.
- GRAUR ALEXANDRU, WALD LUCIA, *Scurtă istorie a lingvisticii*, București, 1965.
- GRAUR AL., *Încercare asupra fondului principal lexical al limbii române*, București, 1954.
- GREIMAS A.J., *La linguistique statistique et la linguistique structurale*, Le français moderne, 4/1962, 1/1963.
- GRUMĂZESCU M., *Studiul asupra foneticii statistice a limbii române*, Comunicările Academiei R.P.R., V, 10, 1955, p. 1429—1437.
- GUIAȘU SILVIU, THEODORESCU RADU, *Matematica și informația*, București, 1965.
- GUIRAUD PIERRE, *Bibliographie critique de la statistique linguistique*, Utrecht-Anvers, 1954.
- GUIRAUD PIERRE, *Problèmes et méthodes de la statistique linguistique*, Dordrecht, Holland, 1959.
- GUIRAUD PIERRE, *La mécanisation de l'analyse quantitative en linguistique*, Etudes de linguistique appliquée, II (Besançon) Paris, 1963.
- HALLE MORRIS, *recenzie la*: Gustav Herdan, *Language as Choice and Chance*, Kratylos, III, 1, 1958, p. 20—28.
- HALLIDAY M.A.K., *recenzie la*: Gustav Herdan, *Language as Choice and Chance*, Archivum Linguisticum, XI, 2, 1959, p. 148—154.
- HARARY FRANK, PAPER HERBERT H., *Toward a General Calculus of Phonemic Distribution*, Language, XXXIII, 2, 1957, p. 143—170.
- HAYDEN REBECCA E., *The Relative Frequency of Phonemes in General-American English*, Word, VI, 1950, 3, p. 217—223.
- HERDAN GUSTAV, *Language as Choice and Chance*, Gromingen, 1956.
- HERDAN GUSTAV, *The Numerical Expression of Selective Variation in the Vowel-Consonant Sequence in English and Russian*, in *Studies Presented to Yoshua Whatmough on His Sixtieth Birthday*, The Hague, 1957.
- HERDAN GUSTAV, *Type-Token Mathematics*, The Hague, 1960.

- HERDAN GUSTAV, *The Calculus of Linguistic Observations*, The Hague, 1962.
- HERDAN GUSTAV, *Statistics of Phonemic Systems*, in *Proceedings of the 4th International Congress of Phonetic Sciences*, Helsinki, 1961, The Hague, 1962, p. 436—439.
- HERDAN GUSTAV, *Quantitative Linguistics*, London, 1964.
- HERDAN GUSTAV, *Quantitative Linguistics or Generative Grammar*, *Linguistics*, 4, 1964, p. 56—65.
- HERDAN GUSTAV, *On Communication between Linguists*, *Linguistics*, 9, 1964, p. 71—76.
- HJELMSLEV L., *The Syllable as a Structural Unit*, in *Proceedings of the Third International Congress of Phonetic Sciences*, Ghent, 1938.
- HOCKETT CH. F., *A Course in Modern Linguistics*, New York, 1958.
- HOOD-ROBERTS A., *A Statistical Linguistics Analysis of American English*, The Hague, 1965.
- IAGLOM I.M., DOBRUSIN R.L., IAGLOM A.M., *Teoria informației și lingvistica*, *Analele româno-sovietice, seria matematică-fizică*, 3 (34), 1960.
- IAGLOM I.M., IAGLOM A.M., *Probabilitate și informație*, București, 1963.
- IONESCU H.M., *Statistica matematică*, București, 1962.
- IORDAN IORGU, *Limba română contemporană*, București, 1956.
- IOSIFESCU MARIUS, MIHOC GHEORGHE, THEODORESCU RADU, *Teoria probabilităților și statistica matematică*, București, 1956.
- IVIĆ MILKA, *Trends in Linguistics*, The Hague, 1965.
- JAKOBSON ROMAN with CHERRY, COLLIN, HALLE, MORRIS, *Toward the Logical Description of Languages in their Phonemic Aspects*, in *Selected Writings*, I, The Hague, 1962, p. 449—463.
- KAISER L., *A Statistical Research of Spoken Language (Dutch)*, *Orbis*, IV, 1, 1955, p. 235—248.
- KLOSTER-JENSEN MARTIN, *Die Silbe in der Phonetik und Phonemik*, *Phonetica*, IX, 1, 1963, p. 17—38.
- KRÁMSKÝ JIŘÍ, *Quantitative Phonemics in the Last Decade*, *Phonetica*, 1—3, 1962, p. 166—185.
- KRÁMSKÝ JIŘÍ, *A Quantitative Phonemic Analysis of Italian Mono-, Di-, and Trisyllabic Words*, *TLP*, 1, 1964, p. 129—143.
- KURILEWICZ J., *Contributions à la théorie de la syllabe*, *Bulletin de la Société Polonaise de linguistique*, VIII, 1948, p. 80—114.
- LEBEDEV C.S., GARMAȘ V.A., *Statisticeskij trebukvennii sočetañij russkogo teksta*, in *Problemi peredaci informații*, 2, 1959, p. 78—80.
- MAČAVARIANI M.V., *On the Relationship between Mathematics and Linguistics*, *Linguistics*, 5, 1964, p. 25—34.
- MACREA D., *Despre fizionomia fonetică a limbii române*, in *Probleme de fonetică*, București, 1953.
- MALMBERG B., *Remarks on a Recent Contribution to the Problem of the Syllable*, *Studia linguistica*, XV, 1961, 1.
- MALMBERG B., *Structural Linguistics and Human Communication*, Berlin-Göttingen-Heidelberg, 1963.
- MANČZAK W., *Fréquence de l'emploi des occlusives labiales, dentales et vélaires*, *BSL de Paris*, LIV, 1959, p. 208—214.
- MARCUS SOLOMON, *Unele semnificații ale lingvisticii matematice*, *LR*, IX, 3, 1960, p. 3—13.
- MARCUS SOLOMON, *Legi cantitative în limbă*, *Gazeta matematică și fizică, seria B*, XI (1960), 4, p. 199—208.

- MARCUS SOLOMON, STATI SORIN, *Entropia limbii*, LR, XII, 6, 1963, p. 581—595.
- MARCUS SOLOMON, *Le modelage mathématique en phonologie. Méthodes, résultats, significations*, CLTA, III, 1966, p. 109—116.
- MARCUS S., NICOLAU ED., STATI S., *Introducere în lingvistică matematică*, București, 1966.
- MARINESCU I., MOINEAGU C., NICULESCU R., RANCU N., URSEANU V., *Elemente de statistică matematică și aplicațiile ei*, București, 1966.
- MATHESIUŠ VILÉM, *Zum Belastungs- und Kombinationsfähigkeit der Phoneme*, in J. Vachek, *A Prague School Reader in Linguistics*, Bloomington, 1964, p. 177—182.
- MATHESIUŠ VILÉM, *La structure phonologique du lexique du Tchèque moderne*, in J. Vachek, *A Prague School Reader in Linguistics*, Bloomington, 1964, p. 156—176.
- MĂRDĂRESCU MARIA, *Frecvența silabelor deschise și închise în limba română literară*, SCL, XVI, 4, 1965, p. 463—467.
- MENZERATH P., MEYER-EPLER W., *Sprachtypologische Untersuchungen*, Studia Linguistica, IV, 1—2, 1950, p. 54—93.
- MILLER GEORGE A., *Langage et communication*, Paris, 1956.
- MOREAU R., *Au sujet de l'utilisation de la notion de fréquence en linguistique*, Cahiers de lexicologie, III, 1961, (Besançon) Paris, 1962.
- NEWMAN EDWIN B., *Statistical Methods in Phonetics*, in L. Kaiser, *Manual of Phonetics*, Amsterdam, 1957, p. 118—126.
- NICOLAU ED., SALA C., ROGERIC AL., *Observații asupra entropiei limbii române*, SCL, X, 1, 1959, p. 35—53.
- NICOLAU EDMOND, *Langage et stratégie*, CLTA, I, 1962, p. 153—177.
- OETTINGER ANTHONY G., *Linguistics and Mathematics*, in *Studies Presented to Joshua Whatmough on His Sixtieth Birthday*, The Hague, 1957.
- ONICESCU O., MIHOC GH., *Lecții de statistică matematică*, București, 1958.
- PETROVICI E., *Esquisse du système phonologique du roumain*, in *For Roman Jakobson, Essays on the Occasion of his 60-th Birthday*, Haga, 1956, p. 382—389.
- PIOTROVSKI R.G., *Problema locului în cuvânt a elementelor purtătoare de informație*, in *Probleme de lingvistică matematică*, București, 1960, p. 32—33.
- PLATH WARREN J., *The Relative Frequency of English Consonantal Patterns*, Z.f. Phon., XI, 1, 1958, p. 67—87.
- PLATH WARREN, *Mathematical Linguistics*, in Christine Mohrman, Alf Sommerfelt, Joshua Whatmough, *Trends in European and American Linguistics, 1930—1960*, Utrecht, 1962.
- REED DAVID W., *A Statistical Approach to Quantitative Linguistic Analysis*, Word, V, 3, 1949, p. 235—247.
- REVZIN I.I., *Models of Language*, London, 1966.
- ROSETTI AL., *Teoria vocabularului* (curs litografiat), București, 1936—1937.
- ROSETTI A., *Filozofia cuvîntului*, București, 1946.
- ROSETTI AL., *Sur la théorie de la syllabe*, The Hague, 1959.
- ROSETTI A., *Remarques sur la définition du „mot”*, CLTA, II, 1965, p. 261—262.
- RUWET NICOLAS, *La linguistique générale aujourd'hui*, Archiv. europ. Sociol., V, 1964, p. 277—310.
- SARAMANDU N., *Sur le rendement fonctionnel des types de structures phonématiques en roumain*, CLTA, III, 1966, p. 147—159.
- SAUSSURE FERDINAND DE, *Cours de linguistique générale*, Paris, 1922.
- SPANG-HANSSEN HENNING, *Typological and Statistical Aspects of Distribution as a Criterion in Linguistic Analysis*, in *Proceedings of the Eighth International Congress of Linguistics*, Oslo, 1959, p. 182—206.

- SPANG-HANSEN HENNING, *Probability and Structural Classification*, Copenhagen, 1959.
- SPANG-HANSEN HENNING, *Code Theory and "Discrete Mathematics" in Phonology*, in *Proceedings of the 5th International Congress of Phonetic Sciences*, Münster, 1964, Basel-New York, 1965.
- STATI SORIN, *La construction des chaînes linguistiques potentielles*, CLTA, II, 1965, p. 267—280.
- X X X *Statistique et analyse linguistique (Colloque de Strasbourg, 20—22 avril 1964)*, Paris, 1964.
- ȘIADBEI I., *Din estetica limbii române: Proporția fonemelor*, Anuarul liceului național din Iași, 1932—1933, p. 12—18.
- ȘUTEU V., *Observații asupra frecvenței cuvintelor în opera unor scriitori români*, SCL, X, 3, 1959, p. 419—443.
- TOCIACU OCTAVIAN, *Unele date statistice privind frecvența literelor și digramelor în limba (serisă) contemporană*, SCL, XVI, 5, 1965, p. 683—722.
- TOPOROV V.N., *Introducerea probabilității în lingvistică și consecințele ei*, în *Probleme de lingvistică matematică*, București, 1960, p. 11—15.
- TROUBETZKOY N.S., *Principes de phonologie*, Paris, 1957.
- UNGEHEUER C., *Einführung in die Informationstheorie unter Berücksichtigung phonetischer Probleme*, Phonetica, 4, 1959, p. 95—106.
- VACHEK JOSEF, *A Prague School Reader in Linguistics*, Bloomington, 1964.
- VACHEK JOSEF, *The Linguistic School of Prague*, Bloomington-London, 1966.
- VASILIU EM., *Une classification des consonnes roumaines d'après le critère de la distribution*, in *Mélanges Oslo*, 1957, p. 97—112.
- VASILIU EM., *Fonologia limbii române*, București, 1965.
- VERTÉS EDIT, *Statistische Untersuchungen über den phonetischen Aufbau der ungarischen Sprache*, Acta Linguistica Ac. Sci. Hung., III, 1953, 1—2, p. 125—158, 3—4, p. 410—430, IV, 1954, 1—2, p. 193—224.
- WEISS MARY, *Über die Relative Häufigkeit der Phoneme des Schwedischen*, SMIL, 1, 1961, p. 41—56.
- ZINDER L.R., *Despre probabilitatea lingvistică*, în *Probleme de lingvistică matematică*, București, 1960, p. 186—195.
- ZIPF GEORGE KINGSLEY, *The Psycho-Biology of Language: an Introduction to Dynamic Philology*, Cambridge Mass., 1965.
- ZWIRNER E., *ZWIRNER K., Die Häufigkeit von Buchstaben und Lautkombinationen*, Forschungen und Fortschritte, XII, 1936, p. 286—287.

PHONOSTATISTICS OF ROMANIAN

The present work is an attempt at a description of the phonological structure of Romanian from a quantitative point of view. From standard Romanian only the aspect 'contemporary Romanian' has been selected.

The method of investigation presented is a rather recent scientific working method and it belongs to modern linguistics. Therefore it has been considered useful to give a series of preliminary remarks having the character of an *introduction* in this field.



1. Until lately 'mathematical linguistics' has been the only general term designating all investigations concerned with the language and having as method mathematics. As a whole, mathematical linguistics is a new discipline. A series of lines of study start from this vast domain and each provides specific images on the language. Out of all disciplines comprised in mathematical linguistics the author dwells especially upon *statistical linguistics*. This is the branch of linguistics which establishes statistical relationships between language elements prevailing at a certain moment. Values are expressed numerically by relative frequencies. The characterization by a certain measure considered relevant is also necessary, besides studying the possible appearance of a certain element in a given situation and the formal relationships between types of structures.

As against other kinds of investigations, statistical linguistics points out facts which represent a *mean* in the language, tendencies, variables and relationships in general, leaving aside the details. It provides information on the general orientation and not on particular events. The structural-qualitative method also investigates

some general aspects of the language, but it does not give any numerical indices. The objective of research renders the statistical method highly useful in the analysis of an ensemble with many elements, hence efficient for the linguistic study. The statistical method makes it possible to detect precisely those regular occurrences with a general character which cannot be detected by other methods. At the same time, it is obvious that quantitative researches on the language do not and cannot suffice for the comprehensive illustration of a material, because of the complexity of the language phenomenon.

The linguistic symbols — phonemes, letters, a.s.o. — occur with a fixed frequency in a given language, since the latter has an ergodic character. This postulate underlying the application of the statistical method is a universally observed and verified truth which may actually be considered as a linguistic law.

Scientific research requires in general *two* types of analysis: a qualitative and a quantitative one. The qualitative analysis observes and classifies the types of phenomena according to their nature, whereas the quantitative analysis (including the statistical analysis) estimates their amplitude and frequency. Obviously, the qualitative appraisal precedes the quantitative one, providing data as to the nature of the elements to be measured. In linguistics both types of research are applied: the method is the same as in the other sciences, only the object differs.

If linguistics is viewed in the ensemble of the other sciences, it is easier to understand that a purely qualitative research is not complete, as the proportion between the elements of the described language becomes obvious only after the application of the quantitative analysis. It must be added that statistics helps linguistics only if it is applied as a working device, subordinate to the main object, i.e. to the qualitative structural analysis of language (even though chronologically elements of linguistic statistics sometimes precede structural linguistics). Therefore, before embarking on any kind of calculation, the linguistic problems to be elucidated must be defined. The determination of the quantitative characteristics of some linguistic phenomena does not mean at any rate their obligatory encompassing within the limits of some mathematical formulae, or their 'deduction' from mathematical structures previously stated.

Observations concerning the history of mathematical linguistics in the framework of its relations with structural linguistics have already been made in speciality literature. An outline of the evolutions of statistical linguistics is presented in this chapter.

The first relation between statistics and linguistics was established a long time ago. According to several authors it might account for the very linguistic systematization of today. The classification of grammatical categories and forms must have been achieved by effecting a selection in the vast material of the languages of the world. Undoubtedly, at first one could not speak of a systematical application of the quantitative method, but only of the use of implicit statistical considerations.

According to some authors, the rigorously scientific statistical linguistics appeared as late as 1940 — with the works of Zipf, Yule, Ross. The quantitative linguistics existing until then cannot be called 'statistical'.

Other authors consider that the first outstanding personality in scientific statistics was Markov, who in 1913 stated the (numerical) laws of form succession in sequences called today 'Markov chains'.

After the publication of E. de Saussure's "Cours de linguistique générale" in 1916, the science of language entered a new stage and the structural schools of linguistics began to be concerned also with questions of quantitative linguistics. The most important moments, the opinions and the works of the members of the Prague School of Linguistics, of the glossematians, and of the American School of Descriptive Linguistics (the Yale School) are surveyed by the author.

At the 6th International Congress of Linguistics (Paris, 1948) the creation of a "Committee for Quantitative Linguistics" was initiated.

The first course on statistical linguistics was held in 1951 in the United States.

Thus, quantitative linguistics emerged from the ensemble of mathematical linguistics as a somewhat independent branch.

At present, international scientific events devoted exclusively to this discipline are organized (e.g. in November 1960 the Scandinavian Symposium of Statistical Linguistics took place at Sundbyholm Manor; in 1964, a Colloquium of Statistics and Linguistic Analysis was held at Strasbourg).



Statistical linguistics — as a distinct discipline — claims to be based theoretically on F. de Saussure's outlook on language. Without asserting that Saussure had a 'quantitative' outlook on language, statistical linguistics selects some principles from descriptive linguistics such as : a) the relationship of interdependence between the sounds and the meaning of a word (the relation *signifiant, signifié*) ; b) the two aspects of language (*langage*), of which 'langue' is the essential element, and speech 'parole' the secondary one. The relation between these two factors may also be formulated as follows : "*langue* is language minus speech".

The object of statistical linguistics — and of linguistics proper — is language as a whole.

The calculus of probability indicates moreover the way in which a language makes use of the material it consists of, the proportion between its component elements.

Statistical linguistics considers language as a system (according to Saussure, "*un système où tout se tient*"), and describes it as a solidarity system between forms pertaining to a certain *level*. The statistical indices of this type of description actually characterize the structure of each linguistic level seen as a whole.



It is no longer necessary to mention that the quantitative method is applied to a series of linguistic fields. However, only those general principles, definitions, etc., which bear directly or indirectly upon the *phonological* system are dealt with. All relations between statistics and linguistics on such levels as the vocabulary, syntax, etc., are not discussed. The term used in the very title of the work, *Phonostatistics*, means "concerning the frequency of phonemes".

The author considered the term '*statistical*' as synonymous with 'quantitative' for the sake of simplicity and because distinctions are optional. A rigorous determination of the fields would lead on to distinctions between statistical aspects, numerical aspects, quantitative aspects. Not every quantitative description is a statistical description (to give just one example, though arithmetic is a quantitative discipline, it does not belong to statistics). At the same time, only questions belonging to the statistical field, that is the facts which have a random character have been followed up.

In general, the object of quantitative phonological typology may be either the phonemes, considered as indivisible units, or

their distinctive features, conceived as structures. The present investigation focusses on the relations between the segmental units, the phonemes; the other aspect (distinctive features) will be discussed in another work after reaching some conclusions on the phonematic entities conceived as phonemes or as groups of phonemes. On the other hand, it is known that any minimal unit set up on the basis of segmentation and phonematic analysis gets linguistic value by the relationships in which it participates necessarily. An aspect is constituted by the relations which prevail on a paradigmatic level. The richer the system, the greater the amount of information borne by each minimal unit. At the same time, as they are elements of a chain of units of expression in a syntagm, between the minimal units reciprocal syntagmatic relations prevail. Syntagmatic structures are important for determining the informative value of phonemes. The distribution rules usually modify the quantity of information per phoneme.

In all languages, the possibilities of building up larger units out of primary entities are limited. As only questions relating to the *connection* of phonemes (with the exception of the chapter where the frequency of each Romanian phoneme is illustrated, in order to achieve an overall image) are to be discussed, the present investigation is carried out on the level of syntagmatic relations. A paradigmatic description would reveal other aspects too.

Phonological elements considered more characteristic for the obtention of an overall image were selected.

Attention was focussed either on the *quantitative relation between units*: vowels/consonants, groups of consonants/groups of vowels (and, within these categories, subdivisions such as: groups of two consonants/groups of three or more consonants, etc.), and on the *relation of transition* (between vowels, between consonants, between vowels and consonants, between consonants and vowels); or on the *position* of the phonematic segments (initial, medial, final); or, finally, on the *size* of the elements under discussion (*how many* units may be comprised in a homogeneous group; what is the length of the most frequent syllabic types; the length of the words, in point of syllables). Also, in such cases the *entropy* obtained in various conditions was calculated. When it has been possible, comparisons with other languages were drawn.

We undertook thus to investigate a series of connected elements significant for the phonological level, and to obtain a series

of information which should not constitute mere "isolated facts of language".

Our work is based on material from written Romanian texts. The sphere of research in the field of the Romanian language is far from being exhausted. For a complete study the obtention of data characteristic of the spoken language will be necessary and will constitute the subject of a future study.

Some authors recommend that investigations should begin with the written language, as the written words illustrate stable, known forms, which are preserved for a longer time in the language.

Moreover the study of written linguistic units is essential for proper understanding in the communication between machine and man and between man and machine.



Before embarking upon the quantitative description of some phonological elements, it is necessary to make some distinctions regarding the notions which will be used in this investigation. A series of terms, such as : frequency, probability, standard error are defined.



For collecting the phonological material a great number of different texts were used. All authors emphasize the importance of the variety and extent of the material for the obtention of relevant quantitative data. Otherwise each series of results would be characteristic only of the texts from which examples are selected. It is only in the case of a great amount of examples that one can speak of the "law of large numbers".

The diversity of the material is ensured by sampling several texts. From mathematical statistics the principle of random sampling was adopted; by samples we understand the group of elements taken from a population and observed by repeated tests. The number of elements in a group is also called the size of the sample.

Generally, literary texts, taken from contemporary prose works were used; in order to have also elements of the current language, less 'stylized' texts from the press (newspapers, reviews) were added. Scientific works which are characterized by a peculiar style, were resorted to only in certain cases.

In general, each example amounted to about ten thousand phonemes. Samples were considered large enough to draw adequate conclusions on the populations presented. In special cases, when we made control probits, checkings, etc., the number of phonemes referred to was mentioned.

The phonological units worked with were phonemes from the inventory set up by E. Vasiliu in *Fonologia limbii române* (The Phonology of Romanian) as this description was also considered more effective statistically.



2. The second chapter of the work includes the description of some types of *quantitative structures*.

Although the Romanian phonemes are of interest especially in determined positions and succession it was considered necessary to know the situation of the phonemes in the language in general, i.e. the quantitative characteristics of phonemes in any position.

The study of the texts reveals the following percent values (with the respective standard errors):

Phoneme	Percent	Standard error percent	Phoneme	Percent	Standard error percent
1	2	3	1	2	3
a	8.8	0.52	l	3.1	0.32
ă	5.7	0.42	m	4.09	0.36
b	0.85	0.14	n	7.6	0.48
k	4.3	0.35	o	5.4	0.41
č	1.2	0.20	p	2.3	0.26
d	3.6	0.35	r	7.5	0.48
e	10.9	0.55	s	4.7	0.39
f	1.1	0.17	ș	1.2	0.20
g	0.9	0.14	t	5.42	0.41
ğ	0.13	0.06	ț	1.02	0.17
h	0.2	0.06	u	7.1	0.47
i	8.1	0.50	v	1.5	0.22
î	3.4	0.32	z	0.5	0.10
3	0.13	0.06			



The quantitative characterization of the Romanian phonemes in various positions, i.e. in various phonological contexts is made.

The investigation of the word *initial* phoneme leads to a series of specific results, as in the hierarchical organization of the component elements of the words the initial part holds a highly interesting place as regards the quantity of information contained in the respective word. Simple initial segments were referred to i.e. phonemes that follow immediately after the pause.

The statistical study of the initial reveals different incidence of phonemes in this position as compared to the incidence of phonemes in all positions (initial, medial and final), accounted for especially on a diachronic level, providing a quantity of information specific to the initial position. *

Vowel groups. This description starts from the set of vowels $V = \{a_1, a_2 \dots a_n\}$ where $a_1, a_2, \dots, a_n$ stand for all vowels of Romanian, and from the set of consonants $C = \{b_1, b_2 \dots b_n\}$ where $b_1, b_2, \dots, b_n$ stand for all consonants of Romanian. Possible combinations between the vowels $a_1, a_2 \dots a_n$ in complex segments are followed up.

The *consonant groups* are investigated in a similar manner.

Both vowels and consonants have been investigated from the point of view of their distributional position, specific contexts being set up. By introducing the two types of sequences in matrices, various symmetries have been established: symmetrical vowel and consonant groups (in which the permutation of the elements results in groups specific to the Romanian language) and asymmetrical vowel and consonant groups (in which a single orientation is possible, and permutation does no longer result in groups specific to the Romanian language); symmetrical vowel and consonant contexts are also extent (vowels and consonants are ordered on either sides of the respective sequence according to a certain ordinate 'axis', formed of this very sequence) as well as asymmetrical vowel and consonant contexts.

We have investigated the word *final*, by which we mean only the phoneme preceding the pause, no matter whether the respective phoneme makes up a syllable by itself or is only a constituent element of a syllable or of a morpheme.

The different incidences are accounted for to a great extent by synchronic data.

It is noteworthy that in Romanian the elements added at the end of words, after the root, are more important for the understanding of the relations between words in the sentence.

3. The third chapter of the work discusses the *lengths of phonematic sequences*.

Syllables and *words* have been dwelt with. Special problems have been raised by the phonological delimitation of the syllables. Types of syllables have been established and then classified according to length, organization and capacity of adherence.

In the syllabic structure there is a primary conditioning of the elements: phonematic units are linked not only by succession, but also by certain rules of reciprocal dependence. In their turn, syllables are organized in words which comprise sequences of successive and interdependent syllables. In other words, chains of phonemes make up syllables, and chains of syllables make up polysyllabic words.

The linking of phonemes in syllables, and of syllables in words does not occur only according to certain rules of concatenation, but also according to certain statistical laws; therefore, the phonematic sequences and, further on, the syllabic sequences make up words, functioning as events with conditioned probabilities.

The distribution of the various types of syllable in the words has been studied.

The investigation of each type of word, as well as of each category (of two, respectively three, four, five, six syllables) separately has been determined by certain particularities of various structures. One type of word does not necessarily 'spring' from another; the three-syllable words do not always mean: the structure of a two-syllable word + the structure of a one-syllable word; four-syllable words are also not necessarily deduced from: a three-syllable word + a one-syllable word, a.s.o.

For each type there are specific ways of combining the syllables: one syllable or various syllables condition the syllable or the syllables which come after. There are syllabic sequences which are possible only in a certain category of words, as there are also sequences which are common to various categories. They differ in the structure, order and frequency of the elements that make them up.

The analysis of syllabic sequences as linguistic chains shows that only some of them are *extensible*.

As the length of words raises specific problems which differ from one style to another, a separate research included words belonging to various styles. Besides the concatenation of various *types* of syllables, the length of the word as a whole was also considered.

Attention was focussed on the factors which determine variations of the words in point of length.

The comparison of longer words from the dictionaries with the words of the investigated texts reveals the fact that, in the texts, the lengthening of words is due less to word-building than to inflexion.



4. The study of language as a message leads to the calculation of the amount of information contained in a message.

The amount of information is determined on the basis of the statistically obtained probability that the event under study may occur.

The quantitative aspect of the information actually constitutes the statistical structure of the information. As the standard form under which the information is given is a string of symbols, it is the frequency of the latter (in the case of language: of phonemes, syllables, etc.) and their various possibilities of combination that are of interest.

The measure of the amount of information given by events succeeding with certain probabilities is the *entropy* (H) introduced by Shannon.

Based on the formula used for determining the entropy, the phonematic entropy of Romanian has been calculated, considering the frequencies of occurrence of the Romanian phonemes.

To the entropy calculated for all texts, as a function of all frequencies of the phonemes, remarks referring to the entropy in *certain* cases in Romanian were added.

Thus, for the special case of initial phonemes:

$$H_{(\text{initial})} = 4.56.$$

The great amount of information obtained by considering only the word initial reveals a certain distinctness. For the receiver of a message, it is very important that he should 'decipher' the general meaning from the very beginning. However, suggestive elements in this respect can be provided by the initial only in a language in which distinctness is an organic factor. Such a feature may be considered characteristic of Romanian as compared to other languages in which the meanings become obvious only in the framework of relationships between larger elements.

Knowing the structure of the syllables of Romanian polysyllabic words, we have calculated the entropy as a function of this element. The structure of a syllable represents the formula of distribution of the vowels and consonants within the framework of syllabic *limits*. Inside the syllables, the phonemes behave like dependent symbols and therefore each syllable has been considered as a whole.

The mean amount of information provided by a type of syllable in Romanian is:

$$H_{(\text{syllable})} = 2.6329$$

Knowing the probability of the appearance of syllabic sequences (2—6 syllables), entropy was calculated for these cases too.

5. The last chapter of the work contains some *conclusions*.

The importance of mathematical studies is based on the preciseness of thought and on the correspondence between actual facts in the language and the symbols by which we try to render them. In our case, the latter are figures.

The quantitative method permits the determination of certain properties with a general character which cannot be detected by other methods. At the same time, more precise details concerning some already known data are given. All these evince the systematic character of language.

Moreover, quantification does not make clear directly, immediately, the qualitative elements of the language under study.

The formalization of language with a view to cybernetic treatment reveals some aspects which cannot be formalized. These help us to discover properties which distinguish language from the machine, as the material of a language is characterized by different kind of complexity than the mechanism of a machine.

In our conception, however, a *many-sided* description of a language cannot be done without quantitative elements.

It is known, for instance, that the inventory of Indo-European languages includes labial phonemes, dental phonemes, etc. Most languages of the world — various linguistic families — contain vowels. These elements — and others, which will not be dwelt upon — make them similar. The quantitative study, however, completes our information regarding the closer, more marked community between the Romanian languages, than between the latter and the Germanic languages for instance.

On the other hand, such a research reveals also what is *specific* to each language taken separately, the relation between its constituent elements.

Certainly, as regards the phonemes, the very determination of the phonological inventory is based on the recording of a great number of variants and the selecting of *some* of the units only. But the determination of the *proportion* between the elements which make up a language is to be preferred to their mere enumeration. Quantification adds another dimension to the system, in which the joining of segments in larger sequences is limited by strict rules. The investigation of possible combinations between phonemes and their frequency reveals the very syntagmatic rules of the system.

It is possible that further studies, some of which might be performed with electronic computers, and on a larger linguistic material, may somewhat modify or complete the results obtained by us.

Therefore, this study is aimed primarily at *illustrating* this method for Romanian. Its aim has been the finding of some relations which cannot be expressed but numerically, and which are believed to be *characteristic*, beside other elements, of the phonologic structure of this language.



To conclude, the following phonostatistical characteristics of Romanian may be pointed out:

I. As regards the types of quantitative structures:

— The Romanian vowel system is made up of seven vowels (a, e, i, ă, î, o, u) and twenty consonants (p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, h, ț, č, ĝ, l, r, m, n); the quantitative relation between vowels and consonants is: 49.4/49.74.

— The simple segments (which comprise but one vowel or one consonant) represent 86.21 per cent, while the complex segments make up 12.65 per cent.

Therefore, when listening to Romanian, one does not get the impression of an 'agglomeration' of sounds; on the contrary, its harmonious structure is evident, vowels and consonants following one another in an equilibrated proportion.

Both in the case of vowel groups and in the case of consonant groups, there is a preference for symmetrical structures;

symmetrical vowel groups (94.5 per cent) are overwhelmingly more frequent than asymmetrical ones (5.5 per cent); symmetrical consonant groups (66 per cent) are twice as frequent as asymmetrical ones (33 per cent).

— As regards the transitions between simple segments VC, CV, the transitions C + V are more frequent; their relation to the transitions of the V + C type is 60.2/25.6 (the transitions whose first element is C being thus 2.3 times as frequent as those with V as first element).

— In certain determined positions, the frequency of phonemes varies, as compared to their overall frequency (determined in connection with the total of their occurrences in standard Romanian, irrespective of their positional conditioning): in initial position, the first three ranks are occupied by consonants; in final position, the first three ranks are occupied by vowels. The explanations provided for this phenomenon have been diachronic for the initials and synchronic for the word finals.

— A summarizing table indicates the 'hierarchy' of the Romanian phonemes:

The most frequent simple segments (considered only in relation with the simple segments in all positions) are:

/e, a, i/

The least frequent simple segments (considered according to the same criterion) are:

/h, ȝ, ʒ/

The variation of the frequency values, from the highest to the lowest, is the following:

for the vowels: between 10.9 and 3.4 — i.e. 7.5;

for the consonants: between 7.6 and 0.13 — i.e. 6.47.

This time too, the differences of the values — as regards the vowels, respectively the consonants — are comparable.

The most frequent complex segments with two elements (considered only in relation with the complex segments, in all positions) are:

for symmetrical vowel groups:

/ea, ui, ie/;

for asymmetrical vowel groups:

/ii/;

for symmetrical consonant groups :

/nd, nt/;

for asymmetrical consonant groups :

/st, jt, nt/.

The most frequent complex segments with more than two elements (considered according to the above criteria) are :

for vowel groups : */eau, ioa, uia, eiau/*;

for consonant groups : */ntr, str, mpl, mpr, kspl/*.

Complex segments with low frequency are not mentioned, as this is not an individualizing factor.

Frequency presents maximum values for the mentioned segments; it shows mean values for a series of groups (vowel and consonant groups), and then it decreases abruptly for a considerable number of sequences.

II. Regarding the length of phonematic sequences :

In general, the most frequent syllables are those which are made up of two phonemes (61.5 per cent of the total of syllable types); the least frequent syllables are those which comprise five phonemes each (0.28 per cent). To state it in a more particular manner, the most frequent structures are of the types : V, VC, CV, CVC; the least frequent are of the types : VCC, CCVCC.

— As regards words, the highest frequency is reached by words with one (39.37 per cent) or two syllables (24.11 per cent).

For two-syllable words, the most frequent type is CV—CV; for three-syllable words, CV—CV—CV; for four-syllable words, CV—CV—CV—CV—; for five-syllable words, CV—CV—CV—CV—CV; and for six-syllable words, CV—CV—CV—CV—CV—CV.

III. Regarding the entropy of Romanian, the following values are noted :

$$H_{(\text{phoneme})} = 4.47$$

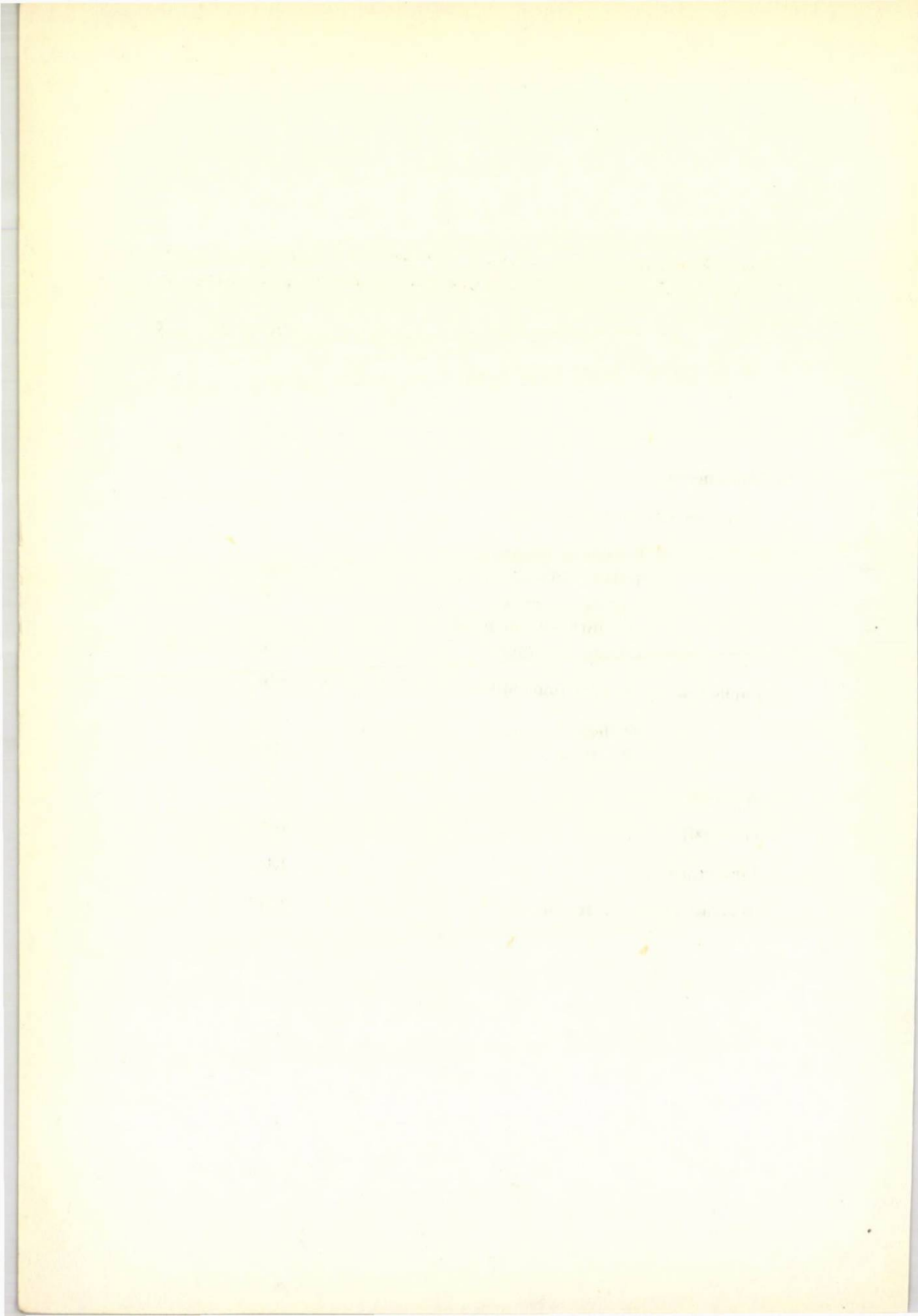
$$H_{(\text{syllable})} = 2.63$$

These data individualize Romanian. At the same time, they represent results which are in agreement with the values obtained for other languages generally comparable with Romanian.



S U M A R

	<u>Pag.</u>
0. Introducere	5
1. Tipuri de structuri cantitative	27
— Fonemele românești	27
— Inițiala cuvintelor	38
— Grupuri vocalice	44
— Grupuri consonantice	63
— Finala cuvintelor	93
2. Lungimea secvențelor fonematice	102
— Silabele	102
— Cuvintele	118
3. Entropia	145
4. Concluzii	150
5. Bibliografie	154
Phonostatistics of Romanian	159



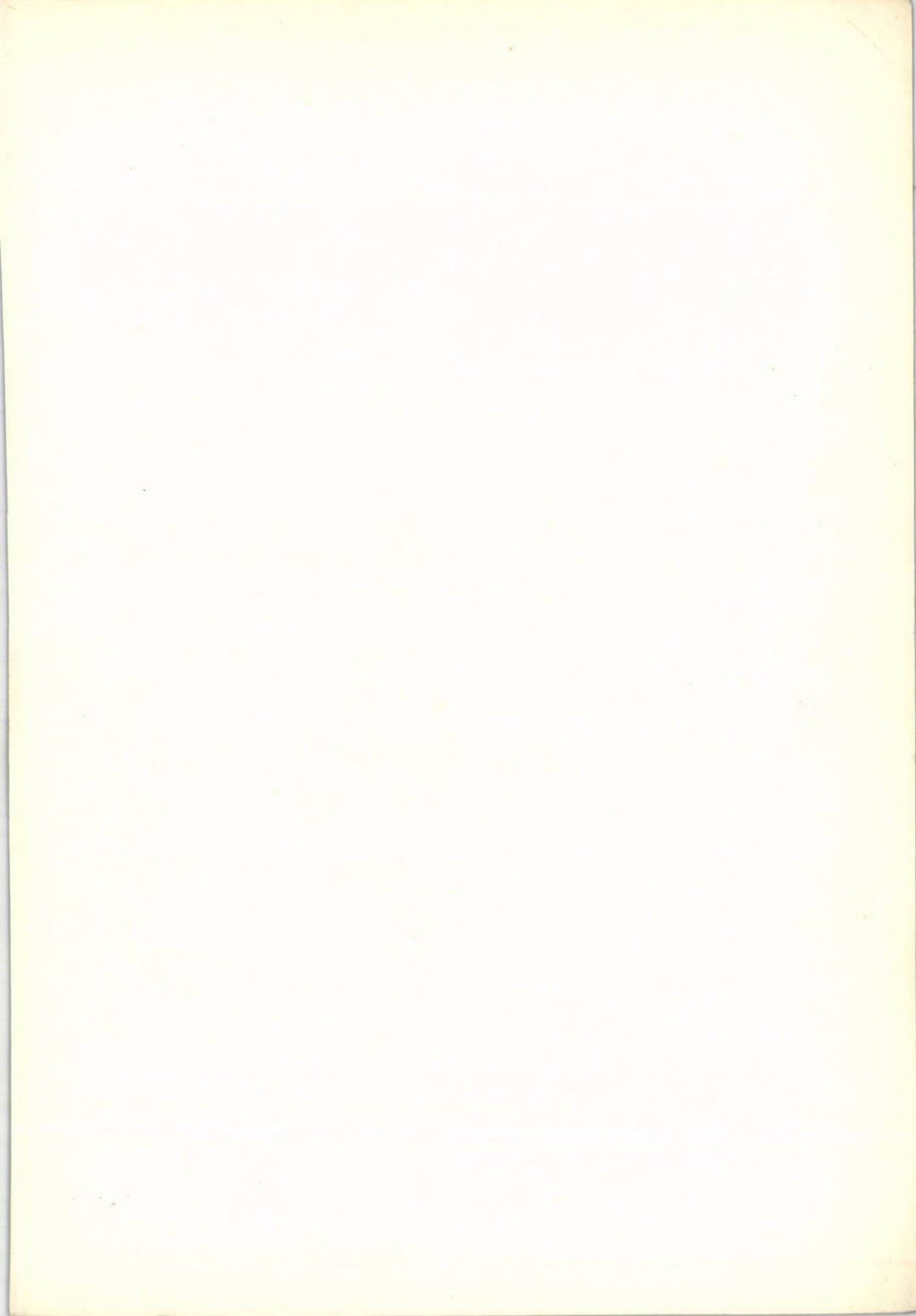
Redactor responsabil: STELLA NICOLAU

Tehnoredactor: PETRU BRUMĂ

*Dal la cules 27.09.1967. Bun de tipar 11.03.1968 apărut 1968.
Tiraj 2440 ex. broșate. Hârtie seris I A de 80 g/m² 610×860/16.
Coli editoriale 10,23. Coli de tipar 11 A 14483/1967 CZ. pentru
bibliotecile mari și mici 459-4.*

Întreprinderea poligrafică „Informația”, Str. Brezoianu nr. 23-25,
București, Republica Socialistă România, comanda nr. 441

36
1



EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA

s
k
p
a
î
d

lei 6